CMOS RF Integrated Circuit Design

Shanghai Short Course

Thomas Lee, Stanford University

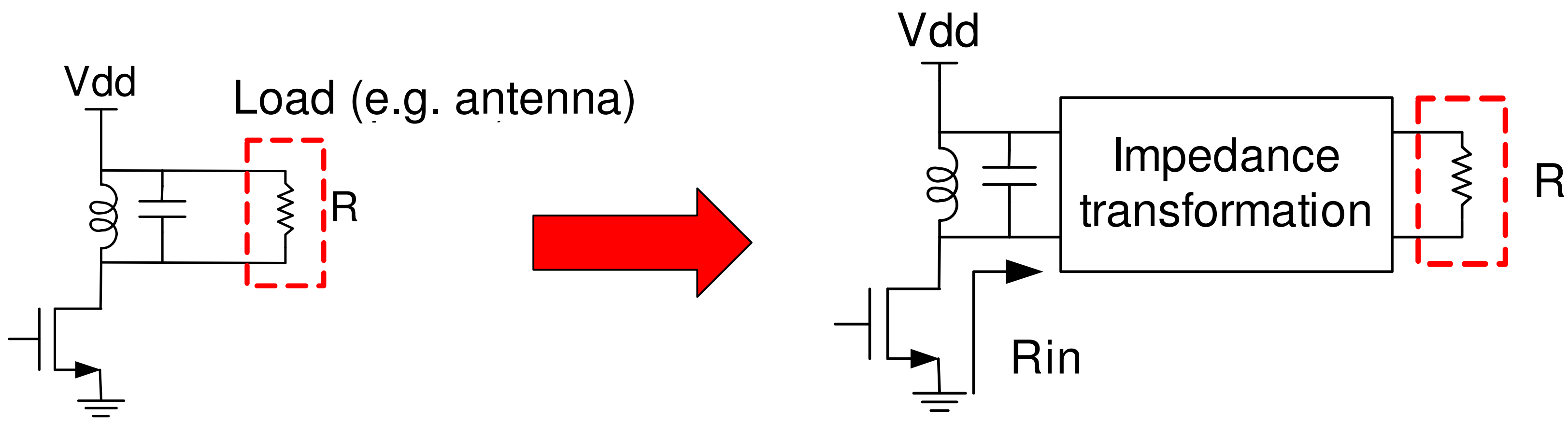
Topics

- Impedance transformation.
- Noise in LTI systems (amplifiers).
- Noise in LTV systems (oscillators and mixers).
- Nonlinearity and what to do about it
- Circuit design at extremes of frequency.
  - When good circuits go bad.
  - Exploitation of parasitics, nonlinearity, and time-variance.

Impedance transformation

- Need to transform impedances arises frequently in RF circuits.
- Some examples:
  - Matching the input of an LNA or PA to 50 ohms (e.g., to avoid reflections on transmission lines).
  - Transforming to a smaller impedance to enable large output power at low supply voltage.
  - Maximizing power gain.
  - Minimizing noise figure.

Why transform impedances?



$$P_{o,\max} = \frac{V_{dd}^2}{2R}$$

$$\left.\begin{aligned} Vdd &= 3V \\ R &= 50\Omega \end{aligned}\right\} \Rightarrow P_{o,\max} = 90mW$$

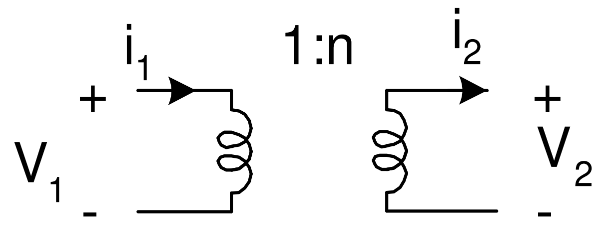
$$P_{o,\max} = \frac{V_{dd}^2}{2R_{in}}$$

$$\left.\begin{aligned} Vdd &= 3V \\ R_{in} &= 5\Omega \end{aligned}\right\} \Rightarrow P_{o,\max} = 900mW$$

- Assuming lossless transformation we can deliver much more power to the load from a given supply voltage.

Textbook ideal transformer

$$\left.\begin{aligned} v_2 &= nv_1 \\ i_2 &= \frac{1}{n} i_1 \end{aligned}\right\} \Rightarrow p_2 = v_2 i_2 = nv_1 \frac{1}{n} i_1 = v_1 i_1 = p_1$$

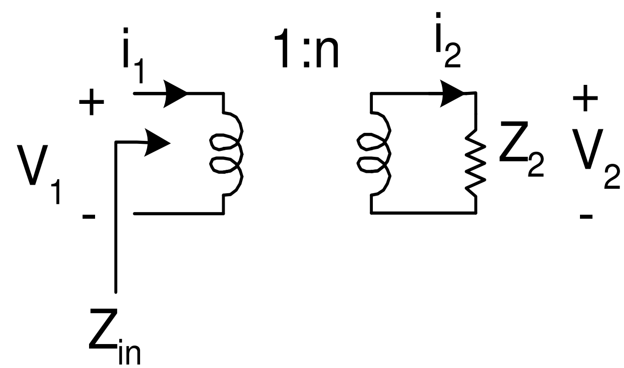


- An ideal transformer has voltage (or current) gain, but no power gain (of course).

$$i_2 = \frac{v_2}{Z_2}$$

$$i_1 = ni_2 = n\frac{v_2}{Z_2} = n^2 \frac{v_1}{Z_2}$$

$$Z_{in} = \frac{v_1}{i_1} = \frac{Z_2}{n^2}$$

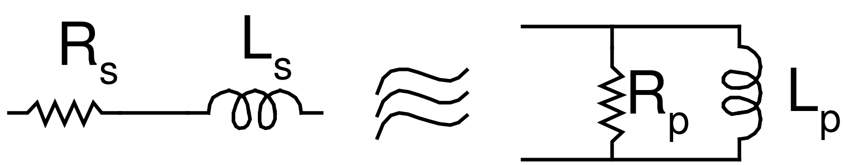
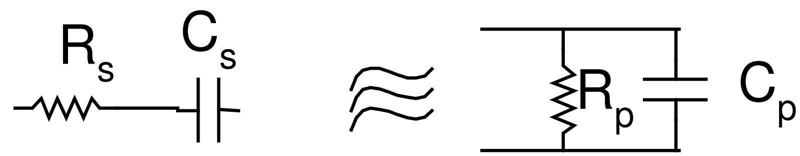


Narrowband impedance transformation

- Textbook transformer is a broadband impedance transformer.
  - Broadband behavior is not always desirable.

- We can replace any series *RC* or *RL* with an equivalent parallel counterpart.
  - A series resistance is transformed to a larger parallel resistance and *vice-versa*.
  - Therefore (in narrowband systems) we can use *LC* networks as impedance transformers.

Equivalent networks (narrowband)

			
$R_p = (1+Q^2)R_s$ $L_p = \frac{(1+Q^2)}{Q^2}L_s$	$Q = \frac{\omega L_s}{R_s} = \frac{R_p}{\omega L_p}$	$R_p = (1+Q^2)R_s$ $C_p = \frac{Q^2}{(1+Q^2)}C_s$	$Q = \frac{1}{\omega R_s C_s}$ $Q = \omega R_p C_p$

- R_p is always larger than R_s.
- X_p is always larger than X_s: $X_p = \frac{(Q^2+1)}{Q^2}X_s = (1+\frac{1}{Q^2})X_s$
  - L_p is larger than L_s.
  - C_p is smaller than C_s.

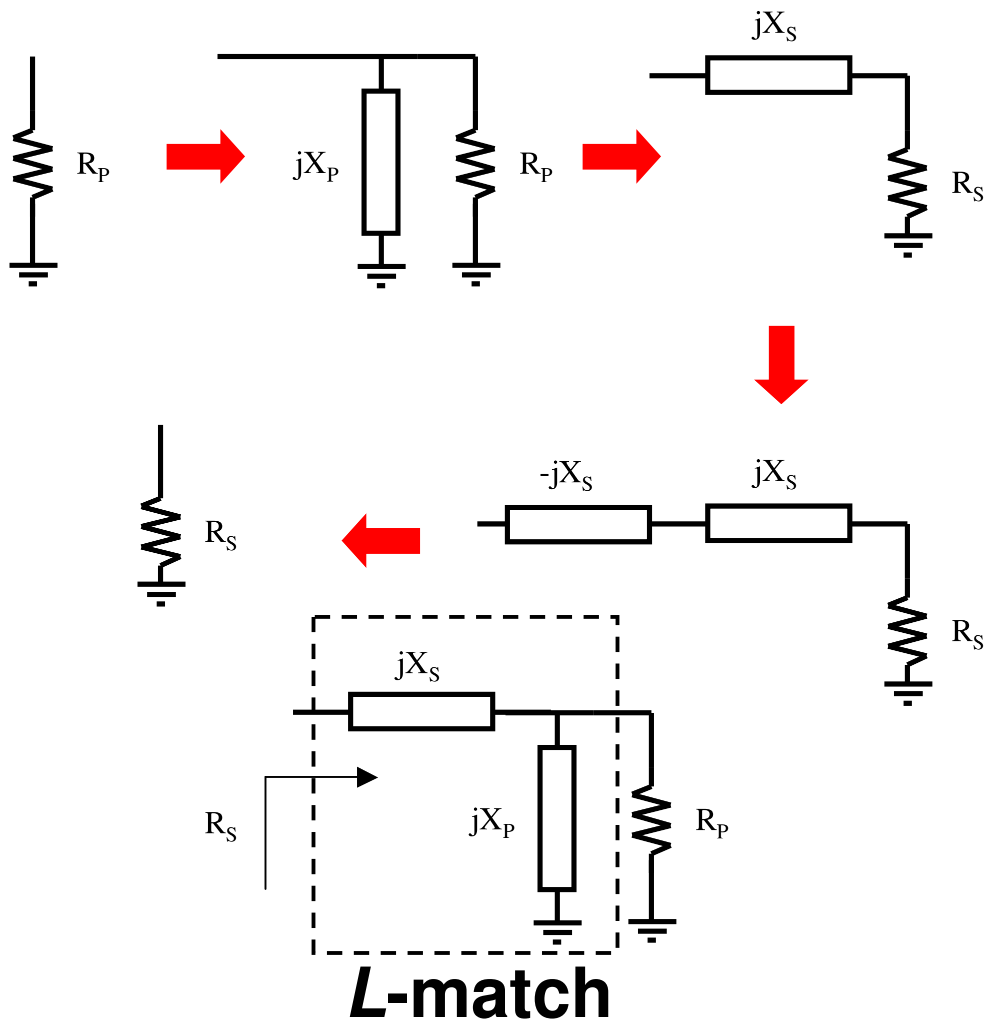
Impedance transformer: The *L*-match

- Suppose we want to transform R_p to a lower impedance R_s

$$1) \quad Q = \sqrt{\frac{R_P}{R_S} - 1}$$

$$2) \quad X_P = \frac{R_P}{Q}$$

$$3) \quad X_S = X_P \left(\frac{Q^2}{Q^2 + 1}\right)$$



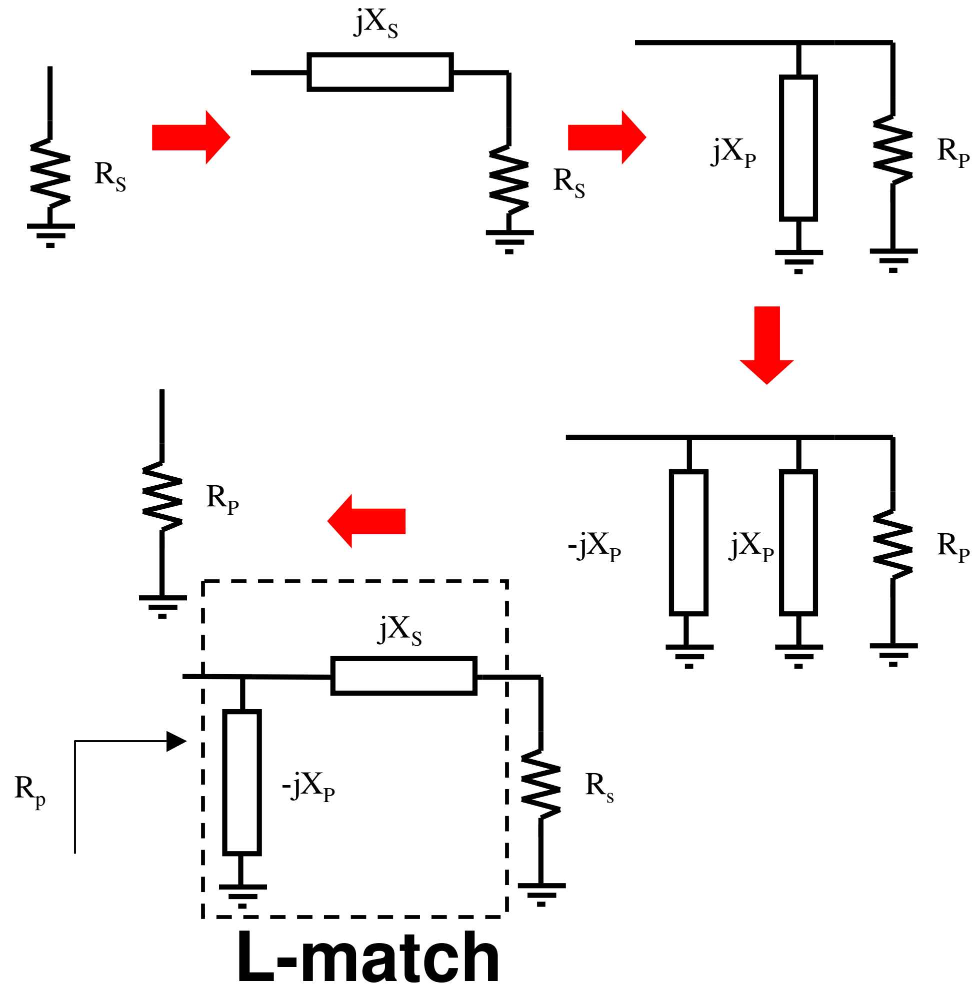
Impedance transformer: The *L*-match

- Now suppose we want to transform R_s to a higher impedance R_p

$$1)\quad Q=\sqrt{\frac{R_P}{R_S}-1}$$

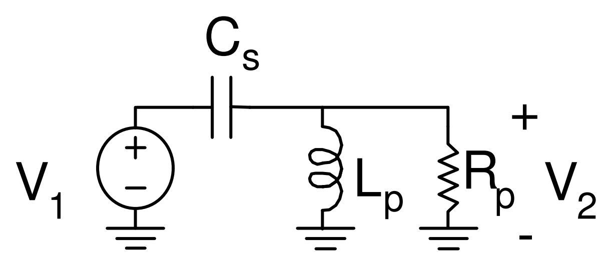
$$2)\quad X_s=QR_s$$

$$3)\quad X_p=X_s\left(\frac{Q^2+1}{Q^2}\right)$$



L-match: Voltage gain

- The *L*-match transforms impedance like a broadband ideal transformer, so it also transforms voltages.



- In a lossless *L*-match all of the power from V_1 is delivered to R_p, so:

$$\frac{V_1^2}{2R_s} = \frac{V_2^2}{2R_p} \Rightarrow \frac{V_2}{V_1} = \sqrt{\frac{R_p}{R_s}} = \sqrt{Q^2+1}$$

- As with an ideal transformer the voltage gain is the square-root of the impedance transformation ratio.

L-match example

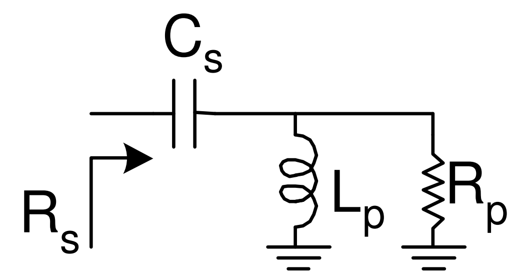
- Ex: Transform 50Ω to 5Ω at 1GHz:

$$Q = \sqrt{\frac{R_p}{R_S} - 1} = 3$$

$$L_p = \frac{R_p}{\omega Q} = \frac{50\Omega}{(2\pi * 10^9)(3)} = 2.6nH$$

$$C_s = \frac{1}{\omega R_s Q} = \frac{1}{(2\pi * 10^9)(5\Omega)(3)} = 10.6pF$$

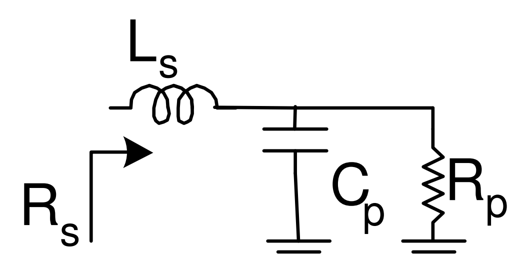
high-pass L-match



$$C_p = \frac{Q}{\omega R_p} = \frac{3}{(2\pi * 10^9)50\Omega} = 9.5pF$$

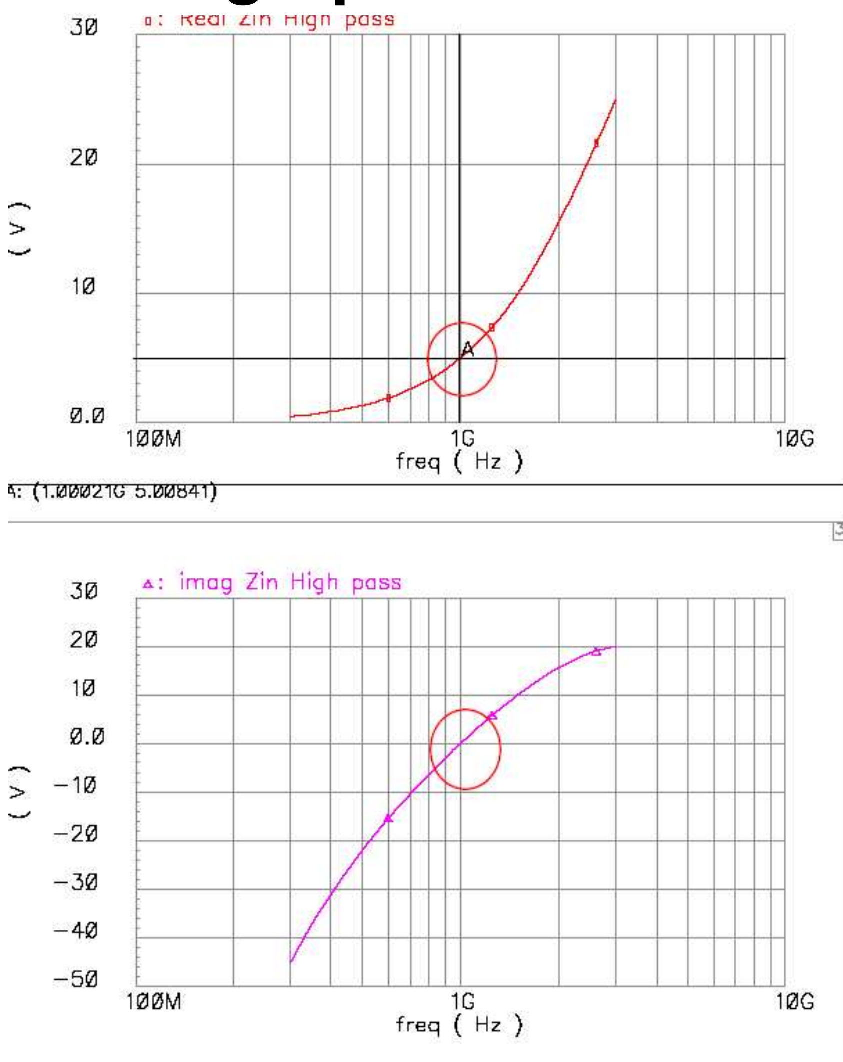
$$L_s = \frac{R_s Q}{\omega} = \frac{3\Omega * 5}{2\pi * 10^9} = 2.4nH$$

low-pass L-match

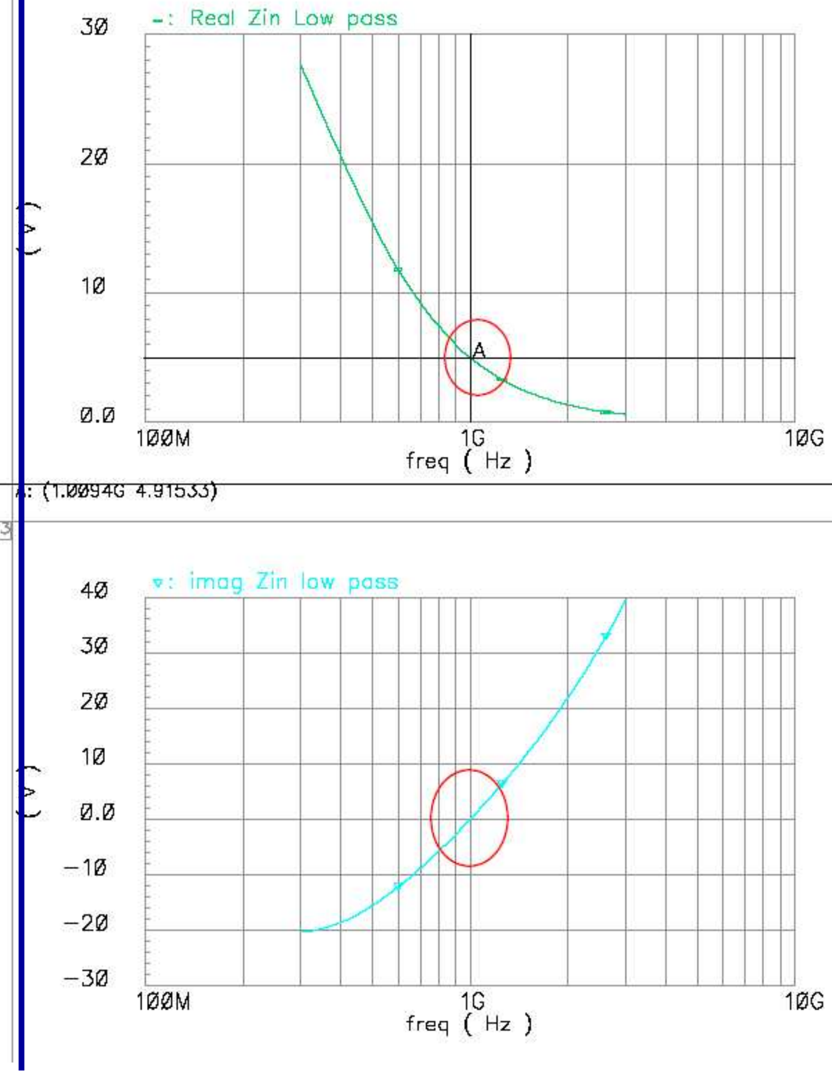


L-match example (cont.)

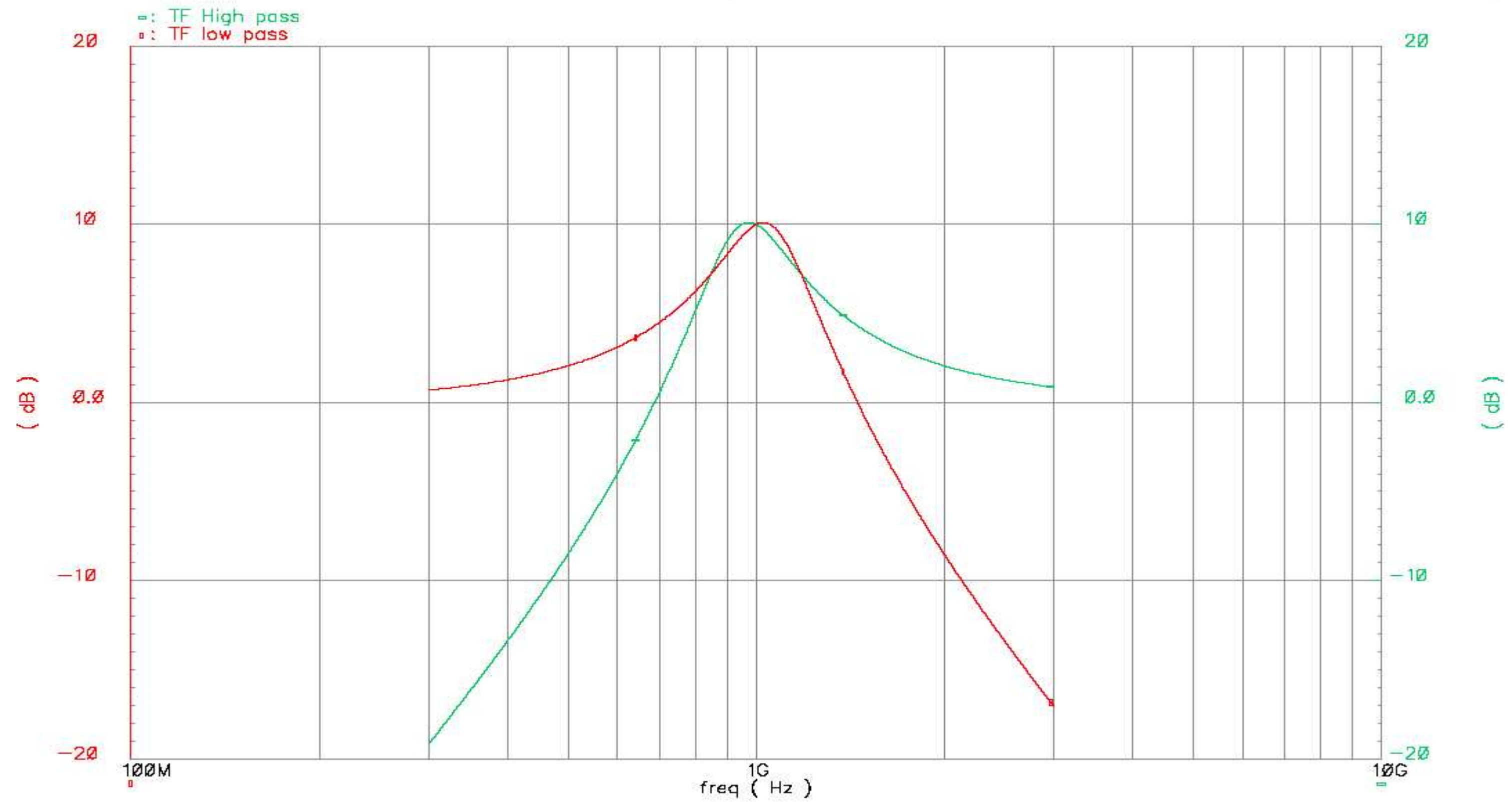
High-pass L-match



Low-pass L-match



L-match example (cont.)



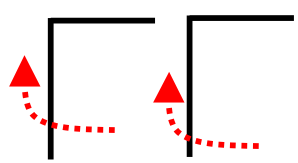
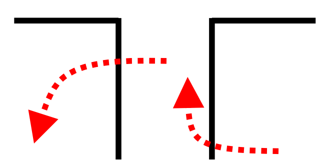
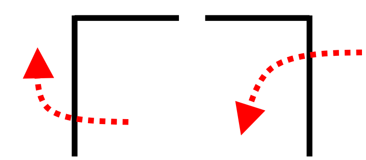
- In the high-pass network voltage gain is unity at high frequencies.
- In the low-pass network voltage gain is unity at DC.

L-match summary

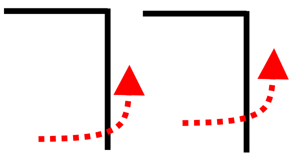
- In a single *L*-match, the matching network's *Q*, and therefore the bandwidth, are set by the transformation ratio.
- Sometimes the bandwidth is unacceptable or the component values are impractical.
- In those cases one would seek additional design degrees of freedom.

Impedance transformers: *T*- and π-match

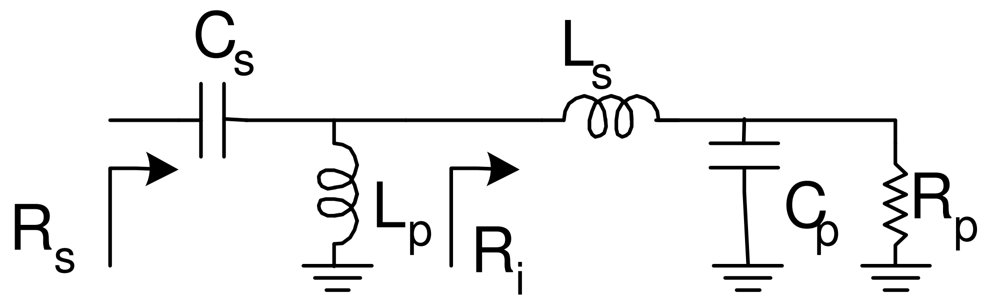
- *L*-match stages can be cascaded for added degrees of freedom.

Cascaded L-match	T-match	π-match
		
▪ "Double-*L*" transforms impedance in only one direction. ▪ Widest bandwidth.	▪ First stage transforms up, and second stage transforms down. ▪ Bandwidth is less than single *L*-match.	▪ First stage transforms down, and second stage transforms up. ▪ Bandwidth is less than single *L*-match.

Cascaded *L*-match

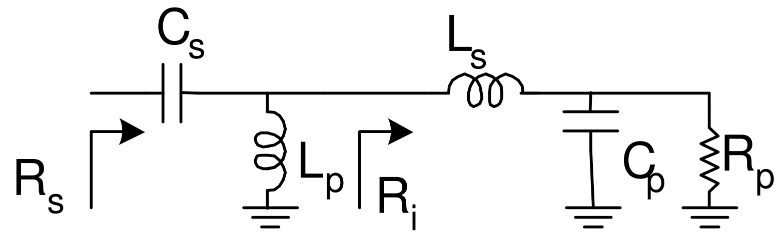


- Specify intermediate impedance R_i to control Q (BW) and response shape. Design problem is then that of two separate *L*-match designs.
- Can combine high-pass and low-pass stages to get both AC coupling and high-frequency (e.g. harmonic) rejection:



Cascaded *L*-match:

(High-pass followed by low-pass)
(R_s=5 , R_p=50, *f*=1GHz)



$$R_i = \sqrt{R_p R_s} = 15.8$$

$$Q_l = \sqrt{\frac{R_i}{R_S} - 1} = Q_r = \sqrt{\frac{R_p}{R_i} - 1} = 1.47$$

$$C_p = \frac{Q}{\omega R_p} = \frac{1.47}{(2\pi * 10^9)50\Omega} = 4.7\,pF$$

$$L_s = \frac{R_i Q}{\omega} = \frac{15.8\Omega * 1.47}{2\pi * 10^9} = 3.70nH$$

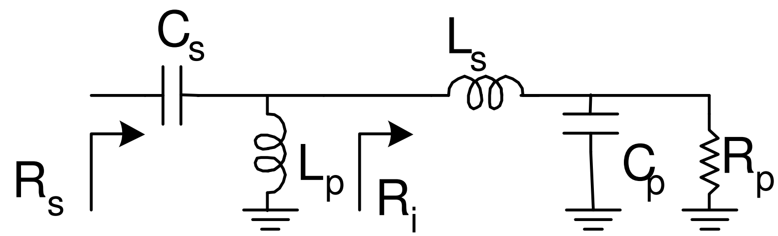
$$L_p = \frac{R_p}{\omega Q} = \frac{15.8\Omega}{(2\pi * 10^9)(1.47)} = 1.7nH$$

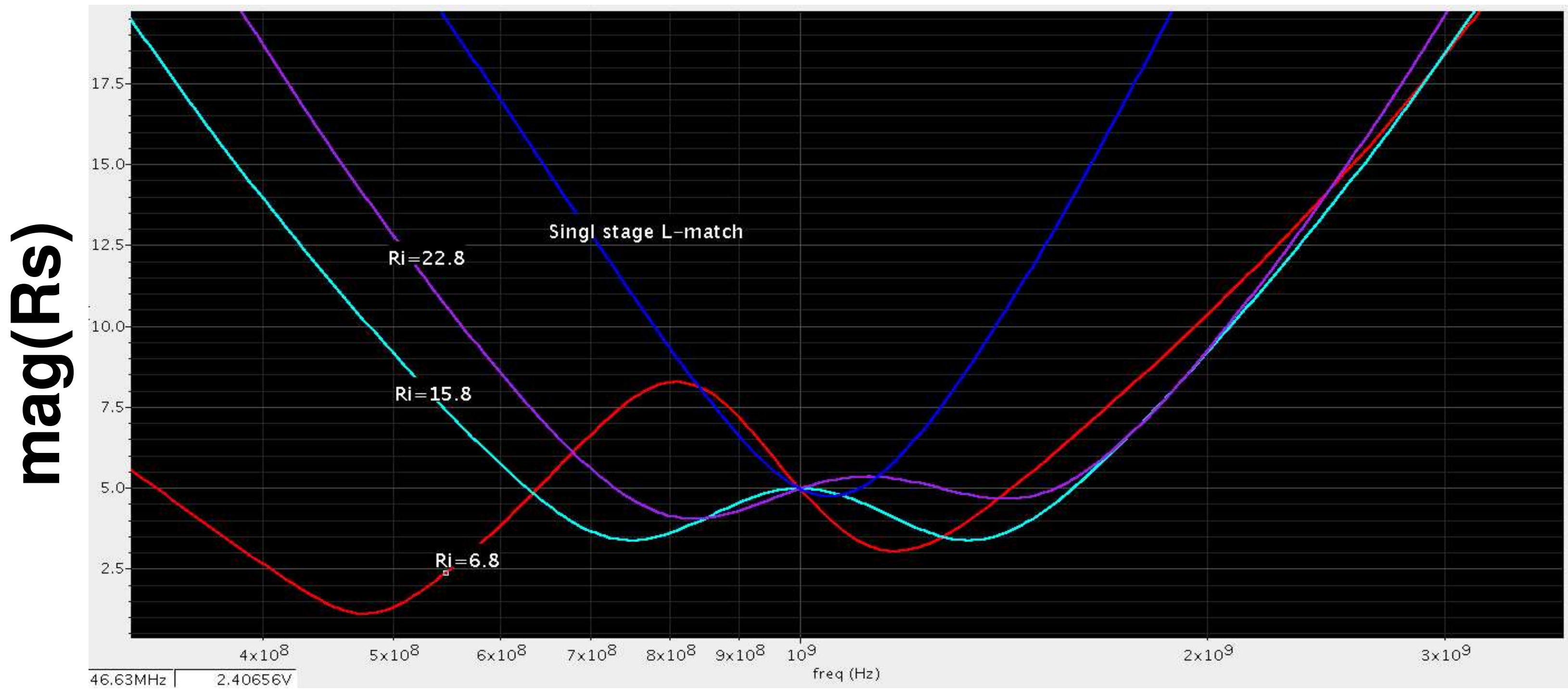
$$C_s = \frac{1}{\omega R_s Q} = \frac{1}{(2\pi * 10^9)(5\Omega)(1.47)} = 21.6\,pF$$

- "Arbitrarily" chose R_i equal to geometric mean of input-output impedances. Component values will differ if different R_i.is chosen.

Cascaded *L*-match

(High-pass followed by low-pass)

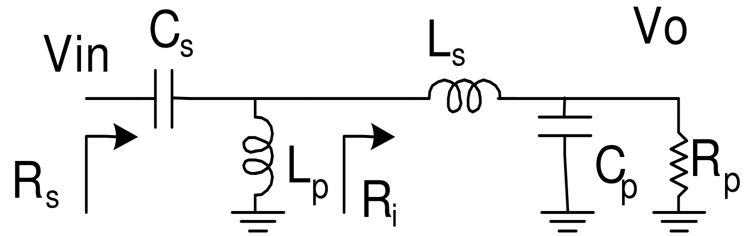




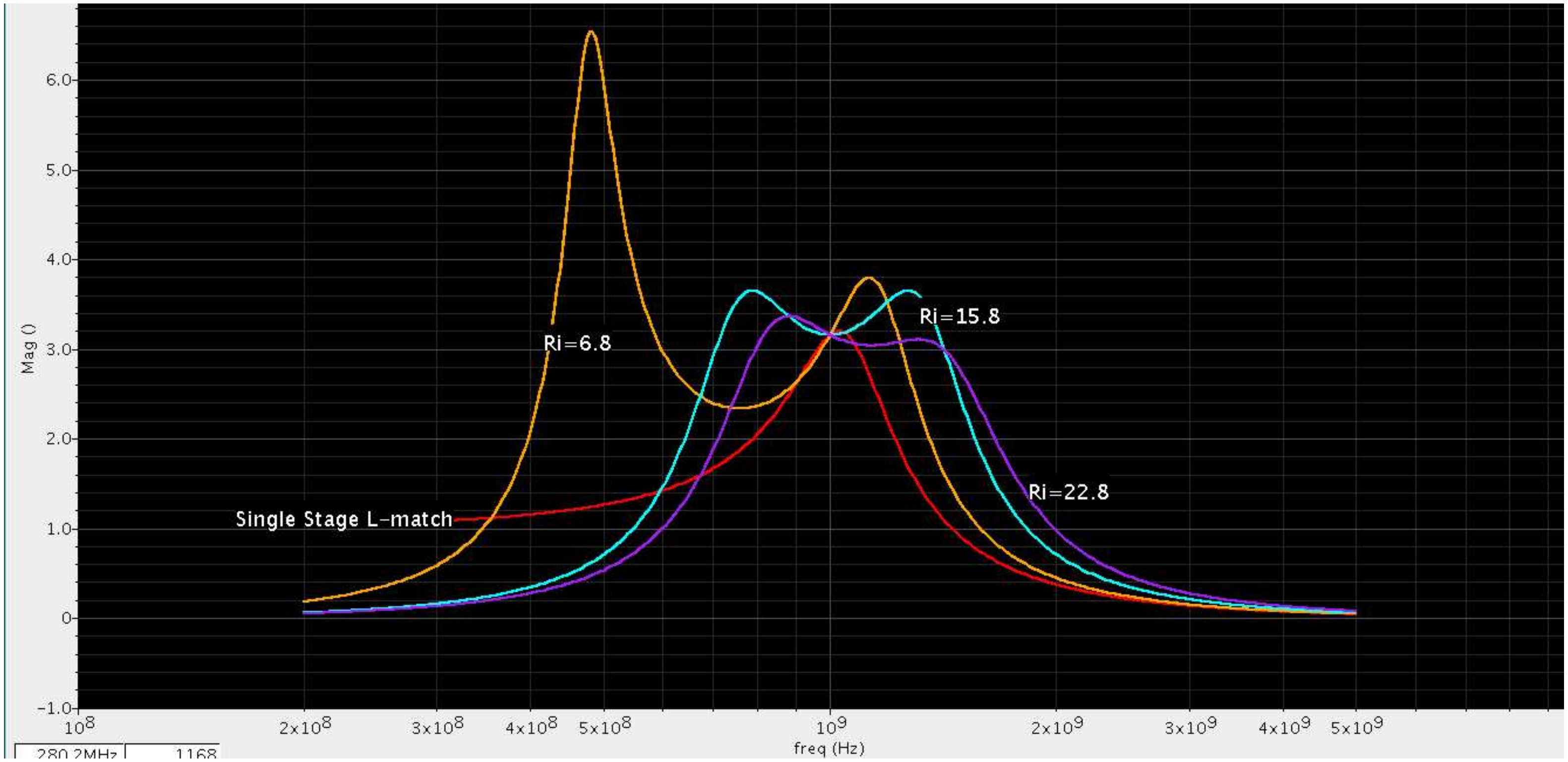
- Symmetric input resistance results when R_i=sqrt(R_p*R_s).
- Symmetric choice also yields widest BW around the design frequency.
- Matching bandwidth is much wider than that of single L-match.

Cascaded *L*-match

(High-pass followed by low-pass)



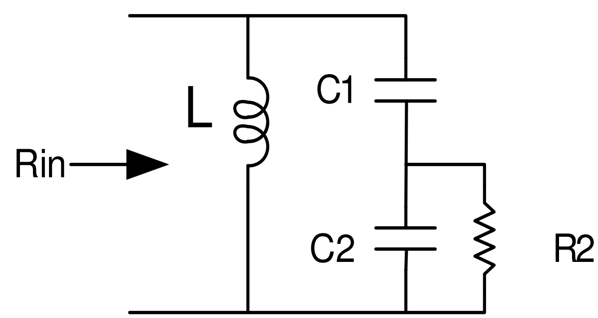
mag(Vo/Vin)



- Symmetric gain results when R_i=sqrt(R_p*R_s).
- R_i=sqrt(R_p*R_s) results in widest BW around the design frequency.
- Gain at design frequency is sqrt(R_p/R_s) = ~3.2 for all cases.

Tapped capacitor network

- Tapped capacitor is just a voltage divider:

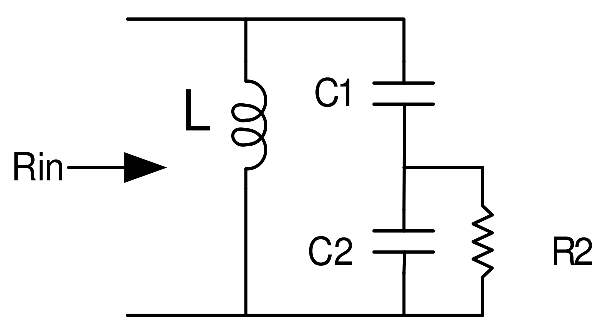


$$\frac{R_2}{R_{in}} \approx \left(\frac{C1}{C1+C2}\right)^2$$

- Impedance transformation ratio is approx. the square of the voltage divide ratio.

Tapped capacitor network

- It's left as an exercise to derive the exact formula. (Hint: Convert R_2-C_2 into series equivalent; combine C_1 and C_{2s} into a single capacitor; convert series RC into parallel RC.)



Why *not* transform impedances?

- Maximizing power transfer isn't always necessary.
  - Impedance transformers consume precious die area, so don't use them unless you have to.
  - Transformers frequently limit bandwidth, too.
- Some amplifiers exhibit improved stability margin with mismatches.
- Later, we will see that most amplifiers exhibit best noise performance with a *mismatch*.

Introduction to noise

Noise

- All electronic circuits generate noise.
- Random thermal (Brownian) motion of charge carriers is ultimately the origin of noise.
- Both passive and active circuits exhibit noise.
- Due to its random nature, noise must be described with the language of statistics.

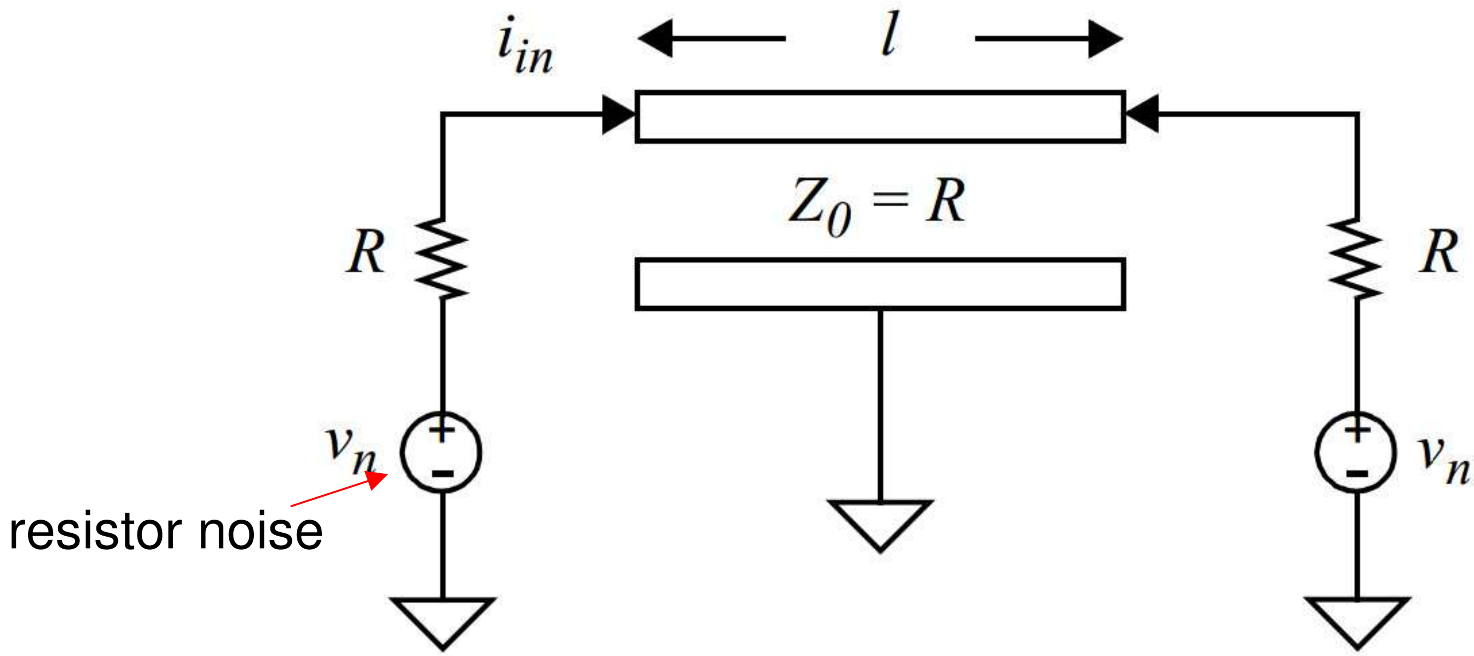
Noise: A little history

- In 1926 Bell Labs' John B. Johnson made careful measurements of thermal noise.
  - Resistors generate noisy voltages/currents!
  - *Q*: Why can't we use this phenomenon as a power source?

- His colleague Harry Nyquist derived the theory with an elegant set of arguments.

Nyquist and noise: "Available power"

- Consider a doubly-terminated T-line:



- At equilibrium, as much power flows to the right as flows to the left.
  - Also true within any frequency interval.

Available power

- Now, the average power supplied by one source to the other is:

$$P_L = \frac{\overline{v_n^2}}{4R} = P_R$$

- Treat noise as random sinusoids propagating back and forth at velocity *v*. Then the average energy stored along a line of length *l* is

$$U = \left(P_L + P_R\right)\frac{l}{v} = \frac{\overline{v_n^2}}{2R}\frac{l}{v}$$

Available power

- If *equipartition* holds, the spectral density is frequency-independent at a value

$$\frac{U}{\Delta f} = \frac{\left(P_L + P_R\right)}{\Delta f}\frac{l}{v} = \frac{\overline{v_n^2}}{2R\Delta f}\frac{l}{v}$$

- Next, compute the spectral density a different way and equate the two expressions.

Available power

- Suddenly short-circuit both ends of the line. The allowed modes have zero voltage at the ends:

$$\lambda_n = \frac{2l}{n} \Rightarrow f_n = n\frac{v}{2l}$$

- If the line is "very long," the distribution of modes will approach a continuum with density

$$\frac{dn}{df} = 2\frac{l}{v}$$

Available power

- Again invoke equipartition, with *kT* energy *per mode* (*kT*/2 each for *E* and *H*):

$$\frac{dn}{df}kT = \frac{2l}{v}kT$$

- Now equate our first expression for energy spectral density to the second:

$$\frac{\overline{v_n^2}}{2R\Delta f}\frac{l}{v} = \frac{2l}{v}kT$$

Available power

- We may therefore write, at last, that

$$\frac{\overline{v_n^2}}{4R} = kT\Delta f$$

- This is the power that would be delivered to a matched load; it's the *available* power, P_{NA}.

- The rhs says that P_{NA} is proportional to the bandwidth of the measurement.
  - And to T.

Thermal noise

- Nyquist showed how thermal agitation of charge results in noisy voltage (and current, via Ohm's law).
  - Amplitude of the thermal noise is Gaussian-distributed with zero mean but non-zero variance (i.e., non-zero power)

$$\overline{v_n(t)} = \frac{1}{T}\int_T v_n(t)dt = 0$$

$$P_{NA} = kT\Delta f \text{ : Available noise power}$$

$$k = 1.38x10^{-23} J/K \text{ : Boltzmann's constant}$$

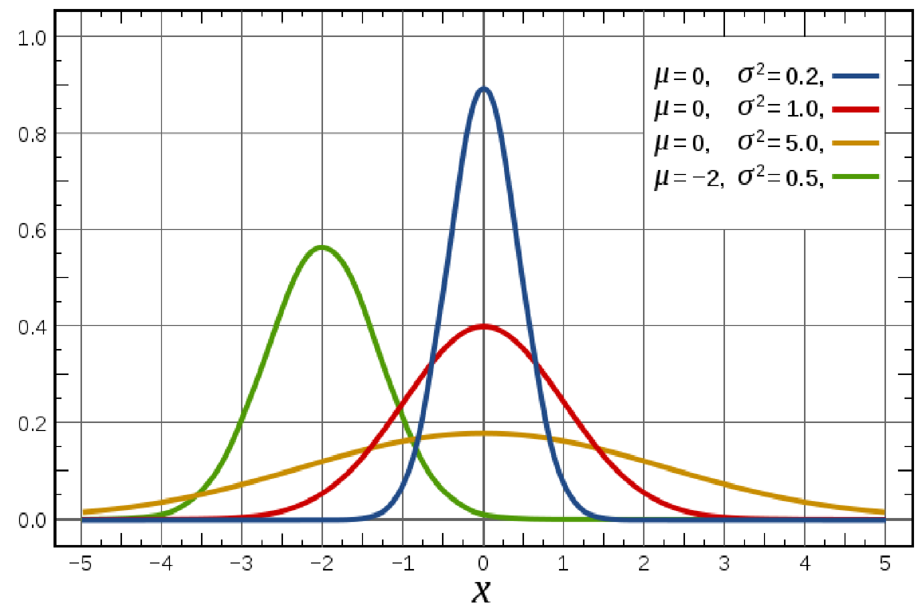
$$T \text{ : Absolute temperature in kelvins}$$

- Remember: Available power is defined as the maximum power that *could* be delivered to a load.
  - It does not have to be delivered; we only care that it is *available*.

Refresher on Gaussian distributions

- The Gaussian (normal) distribution is the most widely used continuous PDF for real-valued random variables.
  - Here, μ is the mean.
  - σ^2 is the variance.
  - If μ=0 then σ is the RMS value.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$$\mu = \int_{-\infty}^{+\infty} x f(x)\,dx :\ \text{1st moment (mean)}$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx :\ \text{2nd moment (variance)}$$

Thermal noise properties

$$P_{NA} = kT\Delta f$$

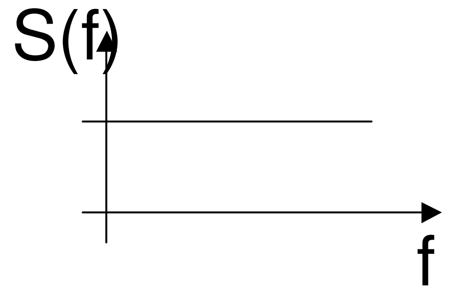
- Available noise power is *independent* of the resistance value.
  - P_{NA}=-174dBm/Hz @ "room" temperature (17°C=290K)

- Thermal noise power is proportional to the absolute temperature.
  - Circuits become noisier when hot!

Thermal noise properties

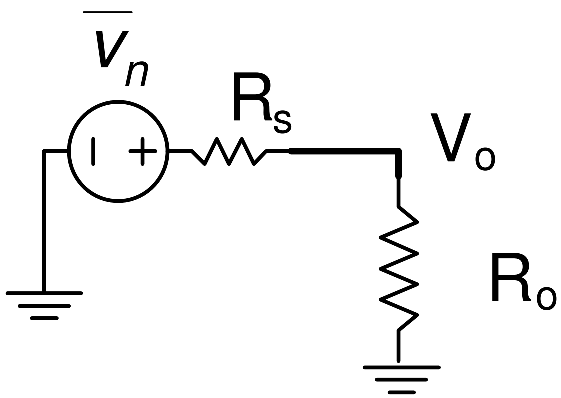
$$P_{NA} = kT\Delta f$$

- Available noise power has flat spectral density $S(f)$ and is thus called "white."
  - Noise power will be infinite if integrated over an infinite bandwidth!



- *Q*: What happens if white noise drives a system with non-flat spectral response?

Resistor noise: P_{NA} vs. rms voltage



$$P_{no} = \frac{V_{no,rms}^{2}}{R_o}$$

$$V_{no,rms}^{2} = \overline{v_n^{2}}\left(\frac{R_0}{R_0 + R}\right)^{2} \xrightarrow{R_0 = R_s} V_{no,rms}^{2} = \frac{\overline{v_n^{2}}}{4}$$

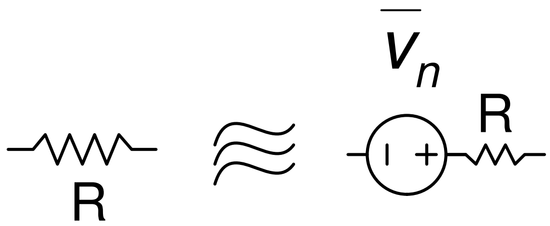
$$P_{NA} = \frac{V_{no,rms}^{2}}{R} = \frac{\overline{v_n^{2}}/4}{R} \Rightarrow \overline{v_n^{2}} = 4kTR\Delta f$$

- $\overline{v_n}$ is the open-circuit rms noise voltage of the resistor.

- Example: If R=1kΩ and Δf=100MHz, then

$$V_{no,rms} = 4nV\sqrt{100MHz} = 40\mu V.$$

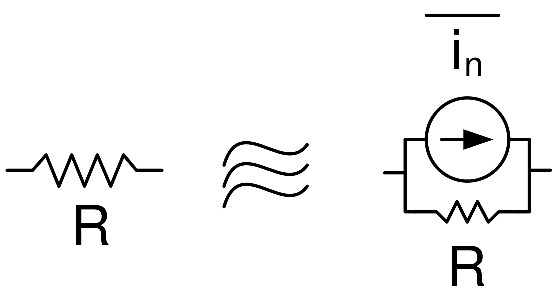
Resistor thermal noise models



$$P_{NA} = kT\Delta f = \frac{\overline{v_n^2}}{4R}$$

$$\overline{v_n^2} = 4kTR\Delta f$$

$\overline{v_n}$ **is the open-circuit rms noise voltage**

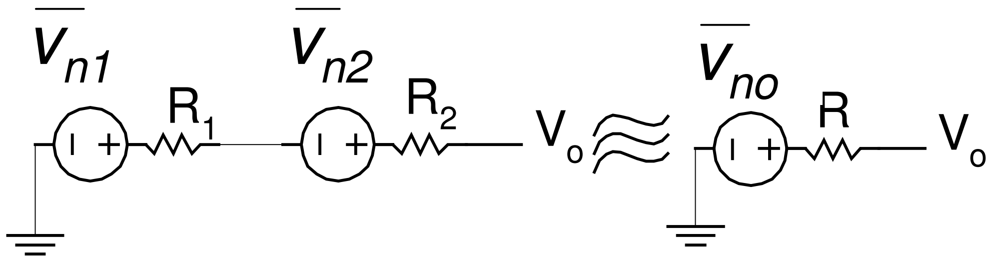


$$P_{NA} = kT\Delta f = \frac{\overline{i_n^2}}{4}R$$

$$\overline{i_n^2} = \frac{4kT}{R}\Delta f$$

$\overline{i_n}$ **is the short-circuit rms noise current**

Series *R* thermal noise (example)



$$R = R_1 + R_2$$

$$v_{no1} = v_{n1}$$

$$v_{no2} = v_{n2}$$

$$v_{no} = v_{no1} + v_{no2}$$

$$\overline{v_{no}^{\ 2}} = E\{(v_{no1} + v_{no2})^2\} = E\{v_{no1}^{\ 2} + v_{no2}^{\ 2} + 2v_{no1}v_{no2}\}$$

$$\overline{v_{no}^{\ 2}} = E\{v_{no1}^{\ 2}\} + E\{v_{no2}^{\ 2}\} + E\{2v_{no1}v_{no2}\} = \overline{v_{no1}^{\ 2}} + \overline{v_{no2}^{\ 2}} + \overbrace{2E\{v_{no1}v_{no1}\}}^{0}$$

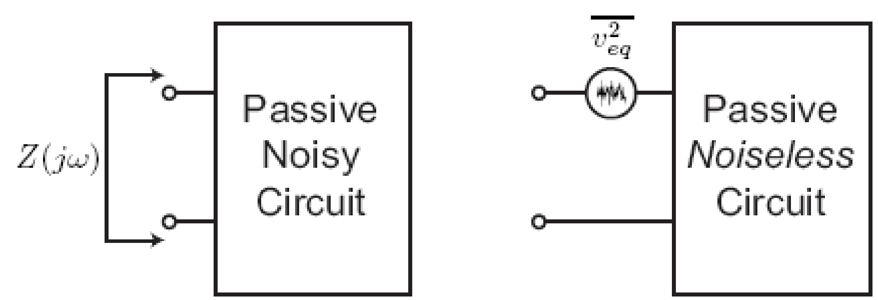
$$\overline{v_{no}^{\ 2}} = \overline{v_{n1}^{\ 2}} + \overline{v_{n2}^{\ 2}} = 4kT(R1 + R2)\Delta f$$

- The noise of R_1 is uncorrelated with that of $R_2 \Rightarrow$ the variance of the output noise voltage (v_{no}) is just the sum of the variances of v_{no1} and v_{no2}.

Thermal noise in *RLC* circuits

- Ideal capacitors and inductors do not *generate* noise, but they may propagate and filter it.
- For a general passive circuit the equivalent mean-squared voltage (or current) noise is:
  - Only a function of the real part of the impedance, not of the imaginary (i.e. reactive) part; noise comes from dissipation.
  - Not white (i.e., NOT flat across frequency). We use ***spot noise*** for voltage (or current) noise at a given frequency.

$$\overline{v^2_{eq}} = 4kT\,\mathrm{Re}\{Z(j2\pi f)\}df$$



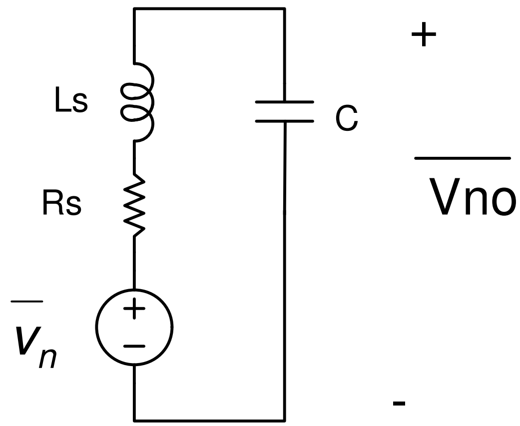
Thermal noise in *RLC* circuits

- To calculate the total noise we integrate the mean-squared noise over the bandwidth of interest:

$$\overline{v^2{}_{eq}} = 4kT\,\mathrm{Re}\{Z(j2\pi f)\}df$$

$$\overline{v^2{}_{eq,T}} = 4kT\int_B \mathrm{Re}\{Z(j2\pi f)\}df$$

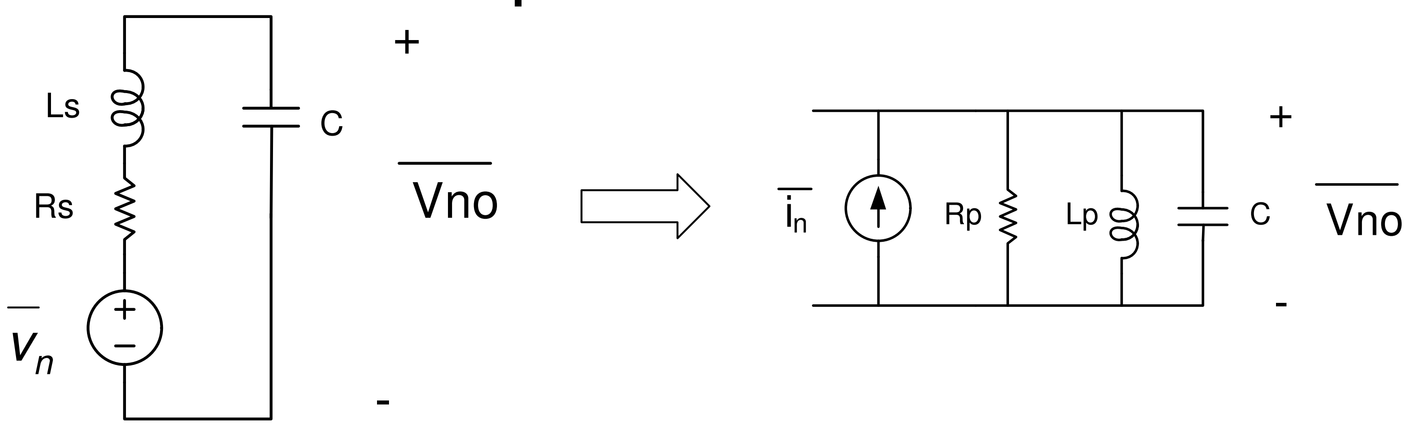
RLC network noise (example)



- Let's consider an *LC* network with lossy *L*.
- What is the rms noise voltage across the *LC* network within a small Δf of the resonant frequency?

RLC noise (example, continued)

- Let's do a series-to-parallel transformation:

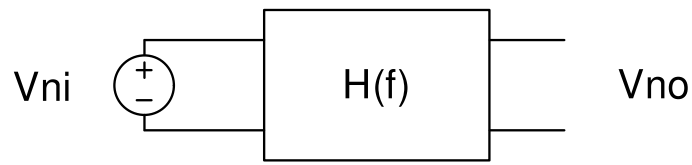


$$Z(jf) \approx \frac{R_p}{1 + j2Q\frac{f}{f_0}} = \frac{R_p}{1 + \left(2Q\frac{f}{f_0}\right)^2}\left(1 - j2Q\frac{f}{f_0}\right)$$

$$-\frac{\Delta f}{2} < f < \frac{\Delta f}{2} \quad \text{and} \quad Q\Delta f << f_0$$

$$\overline{v_{no,rms}^2} = 4kT\int_{-\Delta f/2}^{\Delta f/2}\mathrm{Re}\{Z(jf)\}df = 4kTR_p\int_{-\Delta f/2}^{\Delta f/2}\frac{1}{1+(2Q\frac{f}{f_0})^2}df = 4kTR_p\left.\frac{\tan^{-1}(\frac{2Q}{f_0}f)}{2Q/f_0}\right|_{f=-\Delta f/2}^{f=\Delta f/2} \approx 4kTR_p\Delta f$$

Filtering white noise



$$S_i(f) = kT$$

$$S_o(f) = S_i(f) \times \left|H(f)\right|^2 = kT\left|H(f)\right|^2$$

- The noise spectral density is shaped by the filter so that the output noise spectral density is no longer flat across frequency.
- To calculate the available noise power at the output we need to integrate the noise spectral density over the frequency band:

$$P_{NAo} = \int S_o(f)\,df$$

Equivalent noise bandwidth

- The idea is to define a term for noise bandwidth similar to the way we define bandwidth for signals.
- *Equivalent noise bandwidth* is the bandwidth of a brick-wall (i.e., rectangular) filter with the same peak and total noise power as that of the system in question.

$$\Delta f_n = \frac{1}{\left|H_{pk}\right|^2} \int_0^\infty \left|H(f)\right|^2 df$$

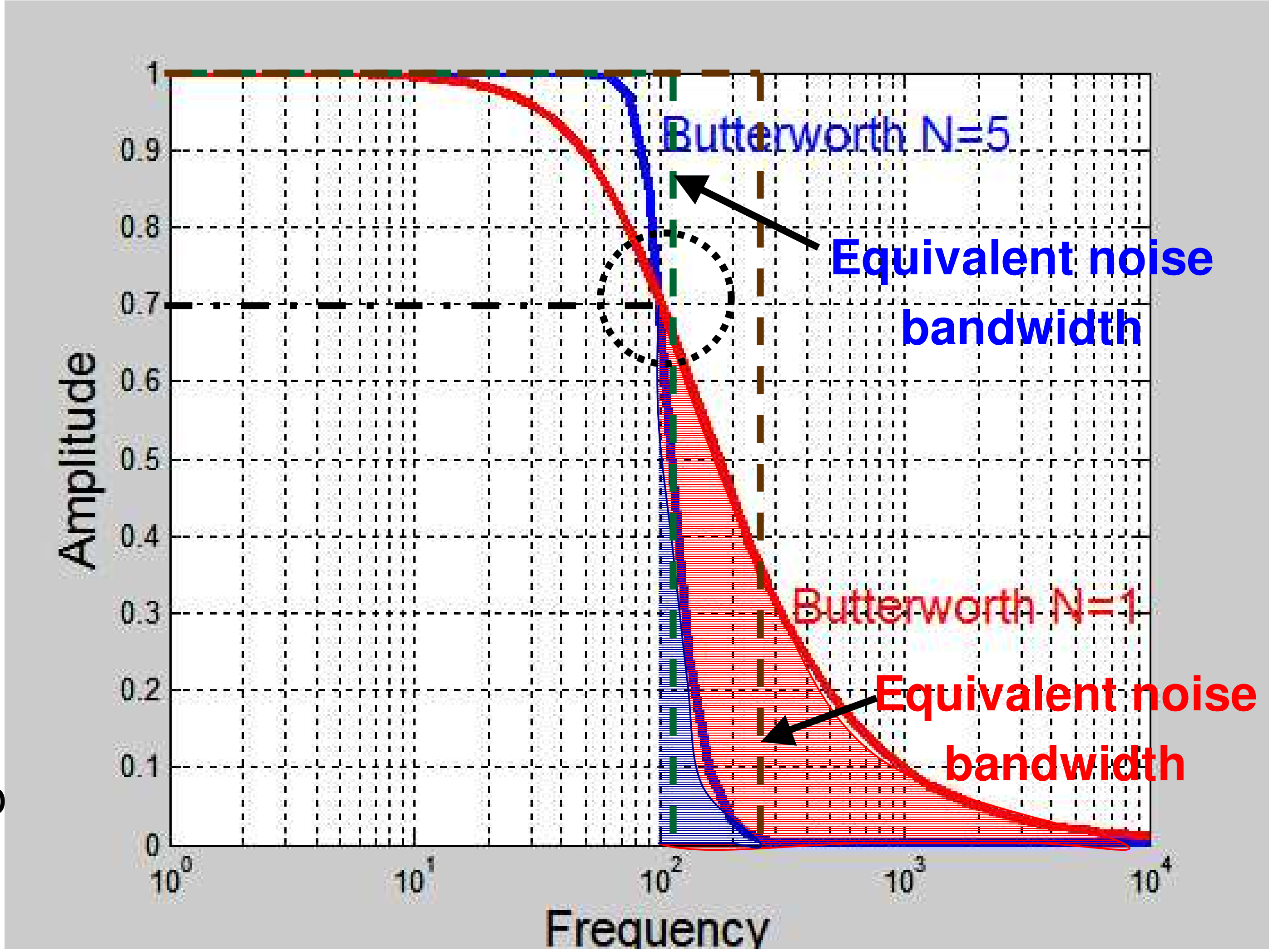
- *Q*: *Is the noise bandwidth the same as -3dB bandwidth*?

Equivalent noise bandwidth (cont.)

- The 3dB bandwidth is simply the half-power frequency.
  - If the rolloff is not abrupt then substantial noise energy will lie beyond the 3dB frequency.
- Let's look at this graphically…

Equivalent noise bandwidth (cont.)

- Consider two Butterworth filters with equal 3dB BW but unequal orders.
- The 5th-order filter's rolloff is steeper than that of the other. Therefore the noise BW of the former is closer to the 3dB BW.



Noise in active devices

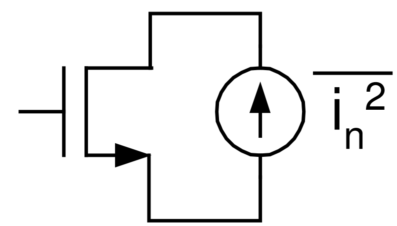
- Nyquist's derivation assumes thermal equilibrium.
- Non-equilibrium is accommodated by adding fudge factors to equilibrium equations.

MOSFET drain current noise

$$\overline{i_n^2} = 4kT\gamma g_{d0}\Delta f\text{; this is a modified resistor current noise eqn.}$$

$$g_{d0}\text{: Drain - source conductance @ } V_{DS} = 0$$

$$\gamma = \begin{cases} 1 & V_{DS} = 0 \\ 2/3 & \text{Saturation (long channel)} \\ \frac{4}{3}\ to\ 2 & \text{Saturation (short channel)} \end{cases}$$



- At low V_{DS} the MOS device behaves like a resistor and γ is unity.
- Excess noise factor (γ) is ~2 times larger in short-channel devices, compared to long-channel devices.

MOSFET drain current noise

- High electric field is partly to blame for the excess noise in short channel devices ⇒
  - Using minimum practical V_{DS} will reduce the excess noise in short-channel devices.
- PMOS devices show less excess noise than NMOS in the short-channel regime.

Flicker noise

- Power spectral density of flicker (1/*f*) noise increases as frequency decreases.

- Sensitivity to surface phenomena (such as charge trapping) worsens 1/*f* noise. MOSFETs exhibit *much* worse 1/*f* noise than bipolar devices.

- Resistors exhibit flicker noise which increases with DC drop across the resistor.

$$\overline{v_n^{\,2}} = \frac{K}{f}\frac{R_{\square}^2}{A} V_{DC}^{\,2} \Delta f$$

MOSFET flicker noise

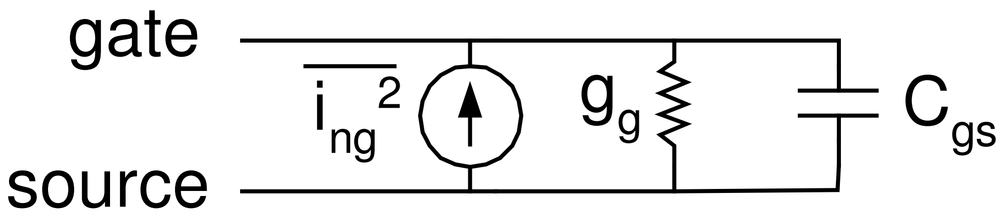
$$\overline{i_n^2} = \frac{K}{f} \frac{g_m^2}{WLC_{ox}^2} \Delta f$$

- The flicker and drain thermal noise are equal at the "1/*f* corner" frequency.
- Increasing the device area (for a given g_m) reduces the 1/*f* corner.

MOSFET gate noise

- Noisy channel charge couples capacitively to the gate and induces a noisy ***gate current***.

$$\overline{i_{ng}^{\,2}} = 4kT\delta g_g \Delta f$$



$$g_g = \frac{\omega^2 C_{gs}^{\,2}}{5 g_{d0}}$$

$$\delta = \frac{4}{3} \approx 2\gamma$$

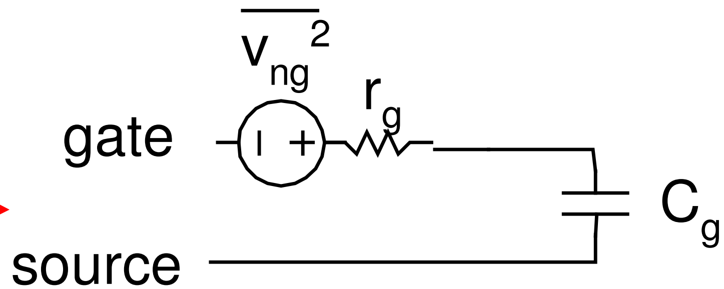
- Gate noise factor (δ) is twice the drain noise factor (γ) in long-channel devices.
  - We often use the same ratio for short-channel devices as well, in the absence of actual data.
  - δ = 8/3-4 for short-channel devices, conservatively.

MOSFET gate noise (cont.)

- Gate noise has the same origin as the drain noise so the two will be *correlated.*
  - ***Cannot simply add their powers to calculate total noise!***
- Gate noise current increases with frequency (it's "*blue,*" not white).
- Gate noise is usually negligible at low frequencies.
  - It can dominate in high frequency/high impedance circuits.
- For those who are bothered by "blue noise," there is an alternative model with white noise:

MOSFET gate noise

- A parallel-to-series transformation works miracles:



ACHTUNG! V_{gs} is defined across the *series combination* of C_g and r_g.

$$\boxed{\overline{v_{ng}^{\;2}} = 4kT\delta r_g \Delta f}$$

$$Q = \frac{\omega C_{gs}}{g_g} = \frac{5 g_{d0}}{\omega C_{gs}} = \frac{5}{\alpha}\frac{\omega_T}{\omega} >> 1$$

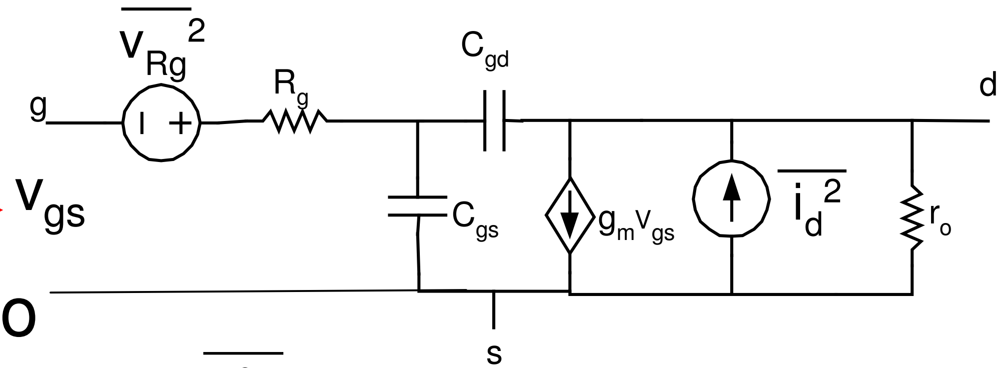
$$C_g = C_{gs}\frac{Q^2+1}{Q^2} \approx C_{gs}$$

$$\boxed{r_g = \frac{1}{g_g}\frac{1}{Q^2+1} \approx \frac{1}{g_g Q^2} = \frac{1}{5 g_{d0}}}$$

- Magically, v_{ng} has a *flat* spectral density!

MOS noise model

Again: V_{gs} is defined across the *series combination* of C_{gs} and R_g.



- Gate resistance has two terms:
  - Induced gate noise resistance.
  - Gate poly resistance.
- *n* is number of gate fingers.
- Factor of 1/3 in r_p comes from the distributed nature of the gate (assuming single-sided contact).

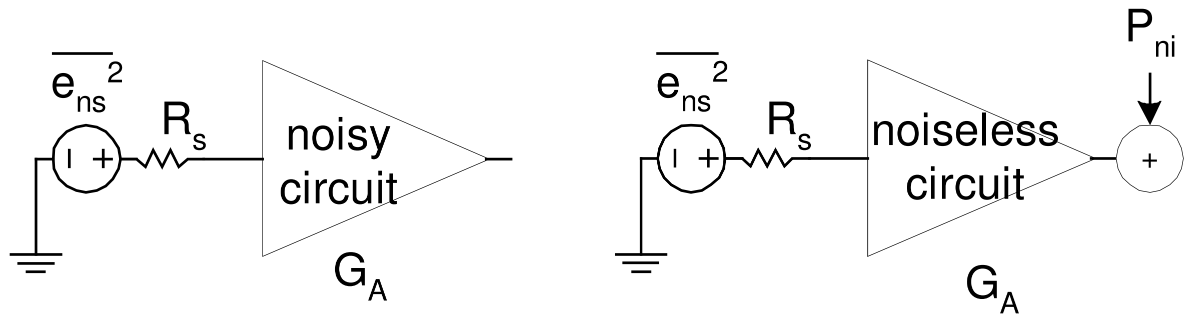
$$\overline{v_{Rg}^2} = \underbrace{4kTr_p\Delta f}_{\text{gate resistance thermal noise}} + \underbrace{4kT\delta r_g\Delta f}_{\text{induced gate noise}}$$

$$r_p = \frac{R_s}{3n^2}\frac{W}{L}$$

$$r_g = \frac{1}{5g_{do}}$$

$$\overline{i_d^2} = \underbrace{4kT\gamma g_{d0}\Delta f}_{\text{thermal noise}} + \underbrace{\frac{K}{f}\frac{g_m^2}{WLC_{ox}^2}\Delta f}_{\text{flicker noise}}$$

Metrics for noise in electronic circuits

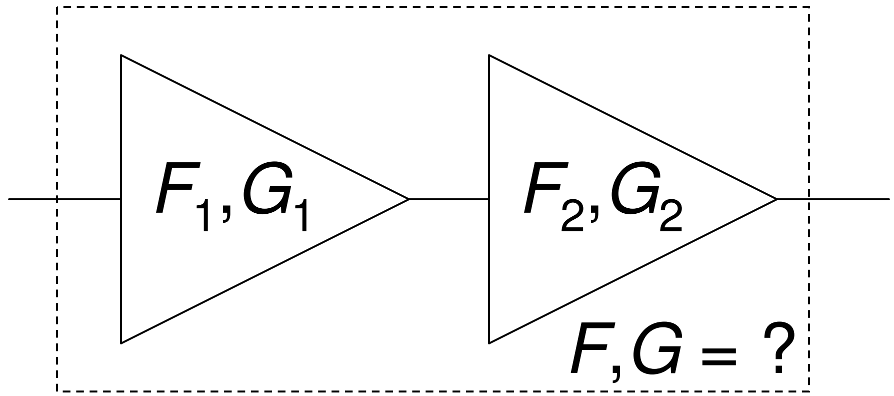


Noise factor (F); evaluated at 290K by convention	$$F = \frac{\text{Total output noise}}{\text{Total output noise due to the source}}$$ $$F = \frac{kTBG_A + P_{ni}}{kTBG_A} = 1 + \frac{P_{ni}}{kTBG_A} \geq 1$$
Noise figure (NF)	$$NF = 10\log(F)$$
Noise temperature (T_N)	$$F = 1 + \frac{T_N}{T_{ref}}$$

Noise figure

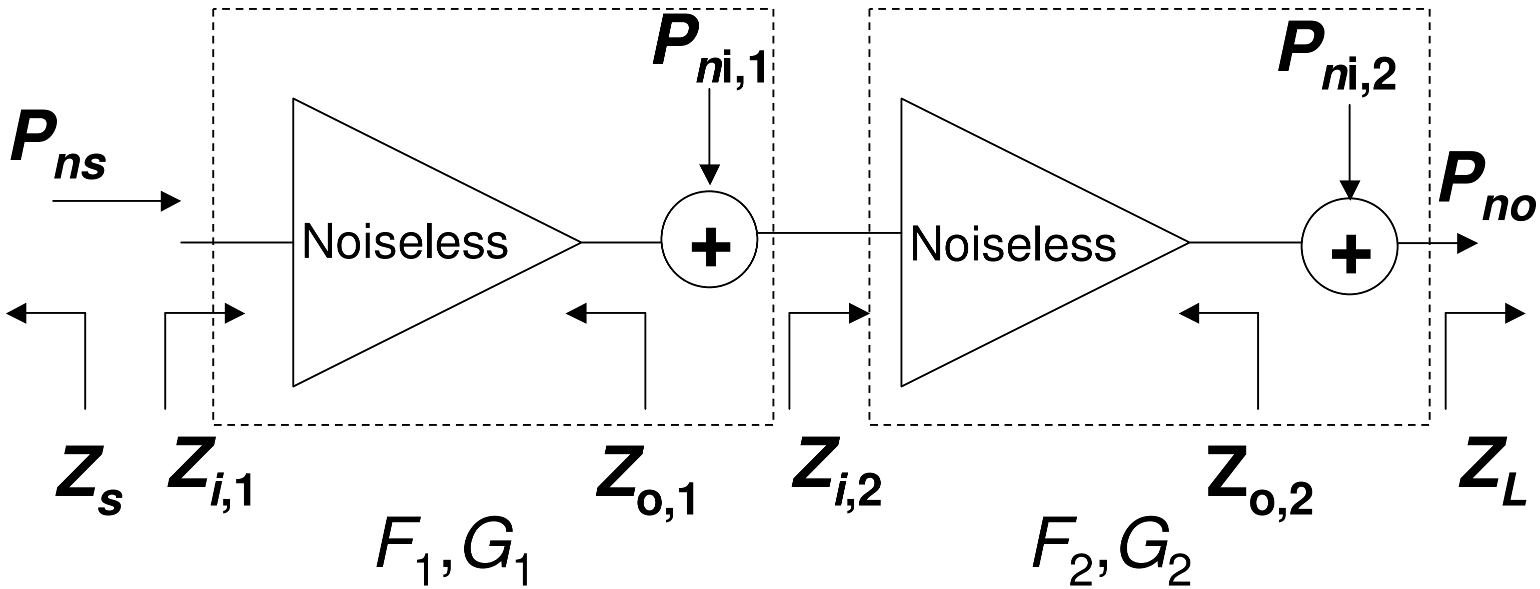
- Noise figures of 2-3dB are generally considered very good.
- Noise figure less than 1dB is outstanding.
- Noise *temperature* is preferred over noise figure when:
  - Noise figure is close to 0dB (noise temperature provides higher resolution).
  - Describing the performance of cascaded amplifiers (the math is simpler, as we'll see).

NF and cascaded amplifiers



- Assume that each input matches the output impedance of the preceding stage.
- What are the gain and noise figure of the overall system?
  - $G = G_1 G_2$ (that's the easy one)

NF calculation in cascaded amplifiers



- Assumption is that all stages are power matched:

$$Z_s = Z_{i,1}^*$$

$$Z_{o,1} = Z_{i,2}^*$$

$$Z_{o,2} = Z_L^*$$

NF and cascaded amplifiers (general case)

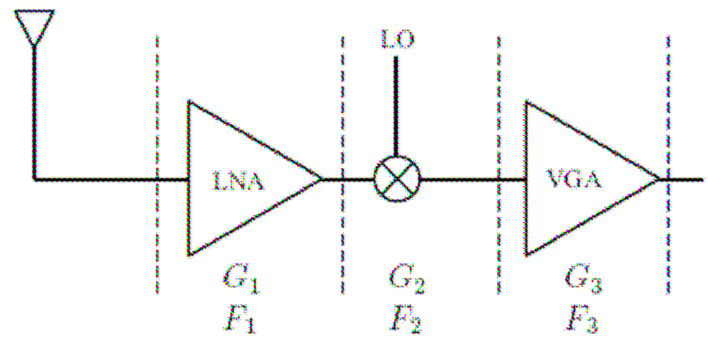
- It can be shown that (Friis' equation):

$$F = 1 + (F_1 - 1) + \frac{(F_2 - 1)}{G_1} + \frac{(F_3 - 1)}{G_1 G_2} + \cdots$$

- If we use noise temperature then the expression is simpler:

$$T_N = T_{N1} + \frac{T_{N2}}{G_1} + \frac{T_{N3}}{G_1 G_2} + \cdots$$

NF (cascade example)

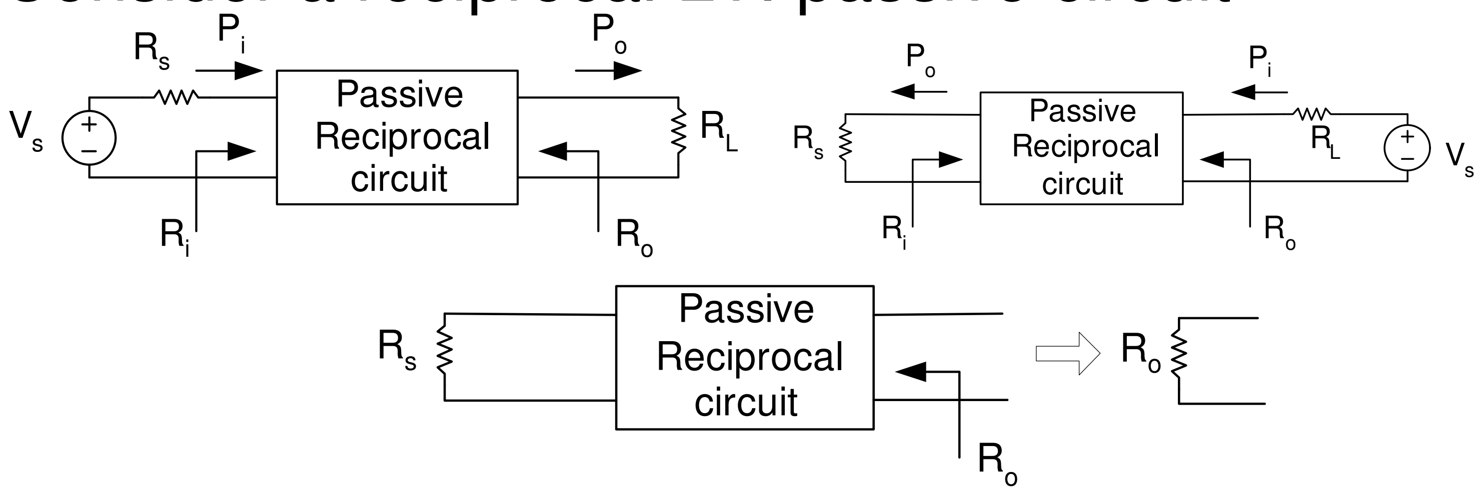


- The LNA has G_1=20dB with NF=2dB. The mixer has a conversion gain of G_2=10dB and NF=12dB. The VGA has max gain of 60dB and NF=20dB.
- As long as the stages are impedance-matched we can apply the cascade formula, even though the circuits operate at different frequencies. Thus,

$$NF = 10\log(1.58 + \frac{15.85-1}{100} + \frac{100-1}{100*10}) = 2.63\text{dB}$$

Lossy reciprocal passive circuits

- Consider a reciprocal LTI passive circuit



- Under an impedance-matched condition we have:

$$R_i = R_s$$

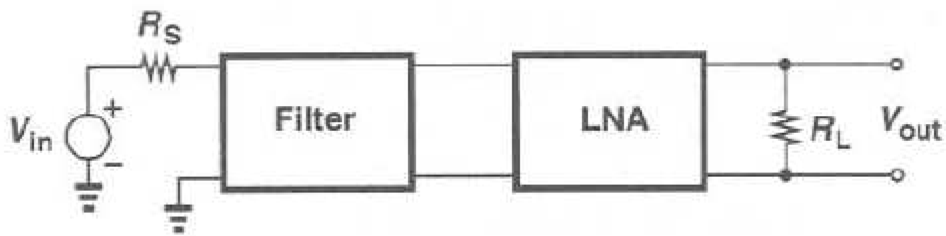
$$R_o = R_L$$

$$L = \frac{P_o}{P_i} < 1 \quad \text{; independent of direction}$$

- It can be shown that $F=1/L$. *"Loss (dissipation) => noise."*

Receiver NF example

- Consider the front-end of a receiver:



- From the Friis equation we have

$$F = 1 + (1/L - 1) + (F_{LNA} - 1)/L = F_{LNA}/L$$

$$NF = 10\log(1/L) + NF_{LNA}$$

- Therefore the loss of the front-end elements degrades the receiver noise figure by precisely that loss, dB for dB.
- Example: If the filter has a 2dB loss the front-end NF will be 2dB+NF_{LNA}

Receiver sensitivity

- Sensitivity is the minimum signal that can be detected with a certain error probability.
- The signal-to-noise-ratio (SNR) is used to calculate the bit-error-rate (BER).
- If the total noise power is N_i, we have:

$$S_{\min}(\text{dB}) = N_i(\text{dB}) + SNR_{\min}(\text{dB})$$

Receiver sensitivity example

- Assume that the receiver requires SNR=10dB to guarantee a BER of 0.1%.
- If NF=10dB and *B*=10MHz, then

$$N_i = P_{ns} + P_{ni} = P_{ns} + P_{ns}(F-1) = P_{ns}F = FkTB$$

$$N_i(dBm) = \underbrace{-174}_{\substack{\text{thermal noise} \\ \text{from source}}} + \underbrace{10\log(10^7)}_{\text{B}} + \underbrace{10}_{NF} = -94\text{dBm}$$

$$S_{\min} = N_i(dBm) + \underbrace{10}_{SNR} = -84\text{dBm}$$

- What should we do if we need -90dBm sensitivity?

Receiver sensitivity example (cont.)

- There is only one solution: Reduce noise!
  - Reducing the bandwidth by 6dB (B_{new}=2.5MHz) would reduce noise, but the bandwidth is often dictated by other constraints.
  - Reducing the noise figure by 6dB is the only practical option.
- Notice that a 6dB increase in sensitivity corresponds to needing a factor of four lower signal power.
- If the received signal strength is proportional to $1/d^2$ (where d=distance), the range doubles.
- Looking at it differently: To have the same range without improving NF the transmitter would have to provide 4x as much power. So, NF is important.

Receiver sensitivity example (cont.)

- To increase the sensitivity by 6dB we should reduce the NF by the same amount, from 10dB to 4dB.
- We are given the option of adding an amplifier in front of the receiver.
- Obviously the amplifier NF must be less than 4dB (why?). If the amp's NF=3dB, then what should be the minimum available gain of the amplifier?

$$F = \underbrace{10^{(4/10)}}_{2.51} = \underbrace{10^{(3/10)}}_{2} + \frac{10-1}{G_{\min}} \Rightarrow G_{\min} = \frac{9}{2.51-2} = 17.65 = 12.5\text{dB}$$

Linearity, Time-Variation, Phase Modulation and Oscillator Phase Noise

Thomas Lee, 17 August 2014

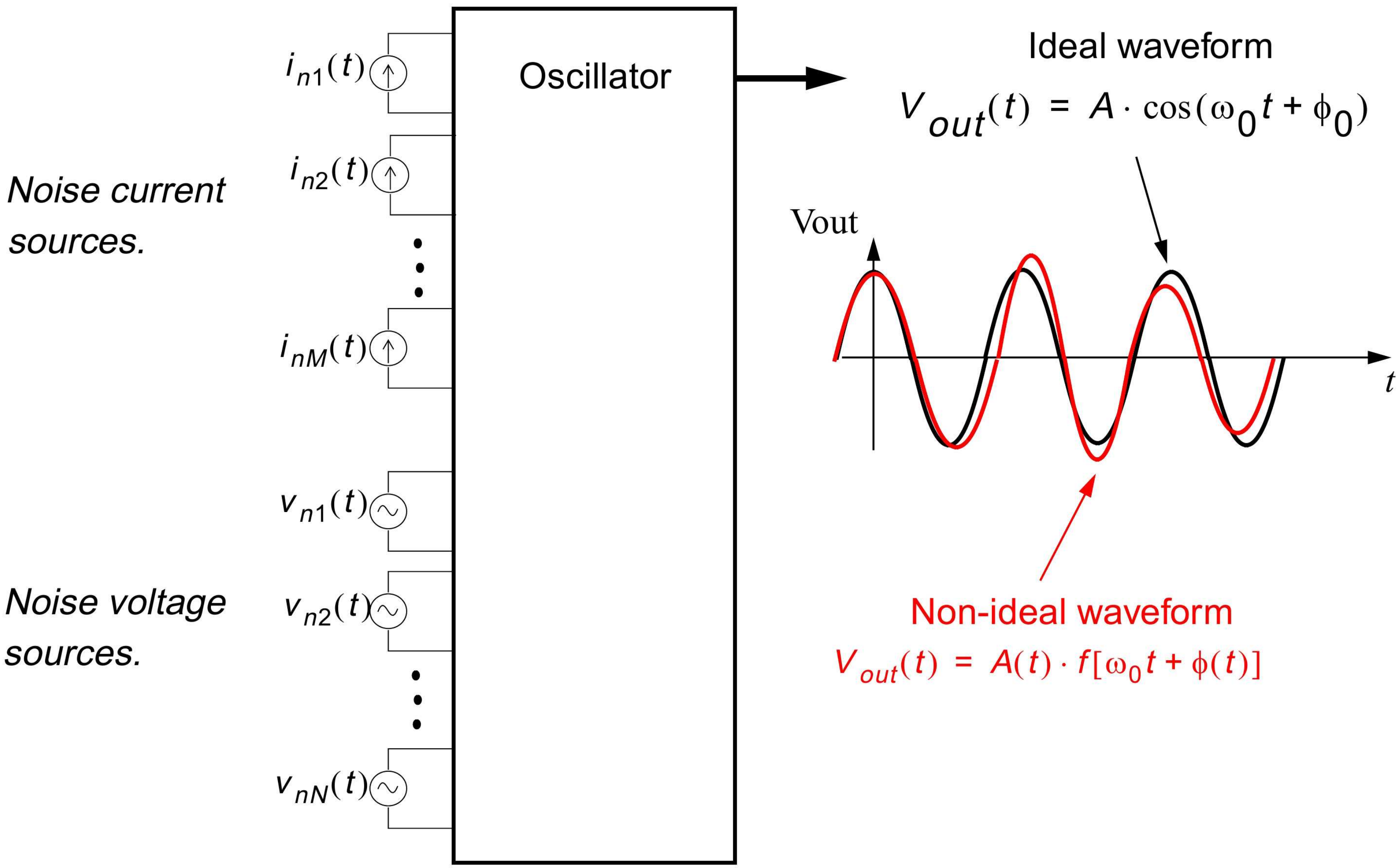
Preliminaries (to refresh dormant neurons)

- A system is linear if superposition holds.
  - Scaling of a single input is included.
  - Impulse response yields sufficient information to deduce the response to an arbitrary input.
  - All real systems can be driven into nonlinearity.
    - Linearity only holds over limited range.
- A system is time-invariant if time-shifting an input *only* time-shifts the output.
- If a system is LTI, then excitation at f produces response only at f.

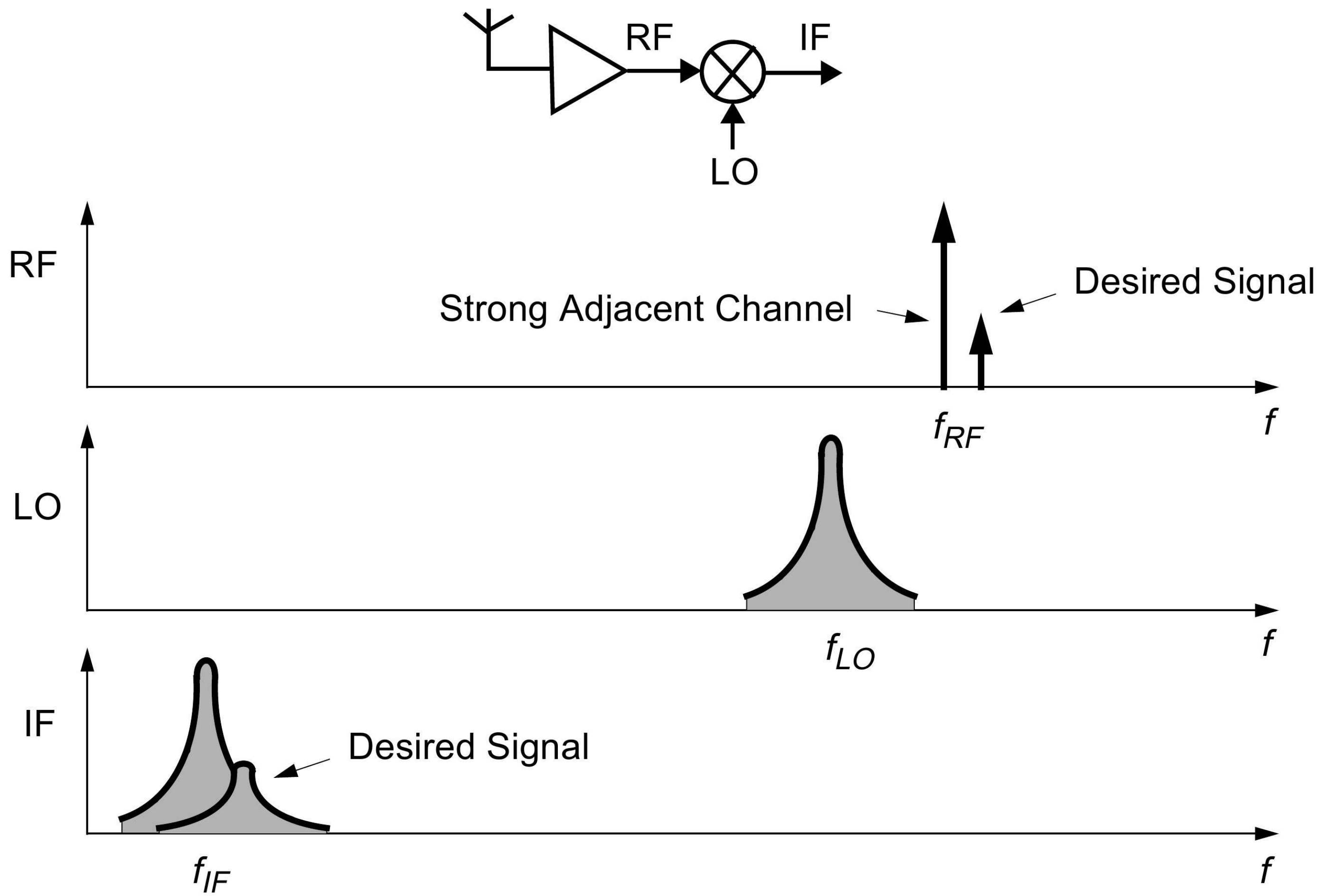
Preliminaries

- If a system is LTV, excitation at *f* can produce response at other than *f*.
  - Superposition holds, so impulse response still tells us about response to any other input.
- If a system is nonlinear, excitation at *f* can also produce response at other than *f*.
  - Superposition doesn't hold; impulse response cannot be used to infer response to arbitrary excitations.

Oscillator as input-output system

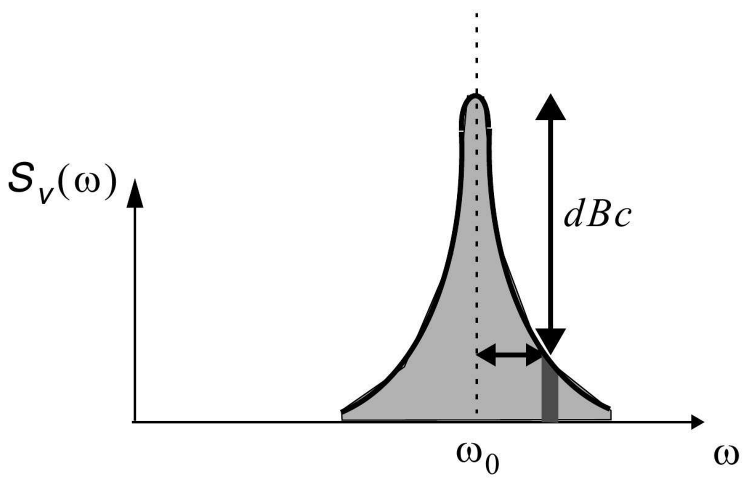


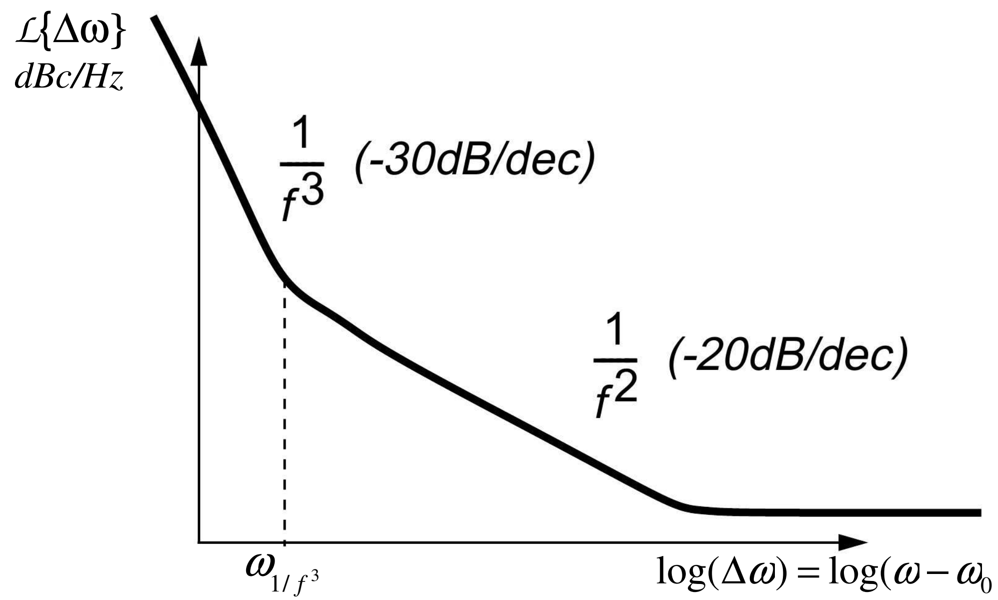
Phase noise effects in RF systems



Phase noise buries desired signal in skirts of interferer

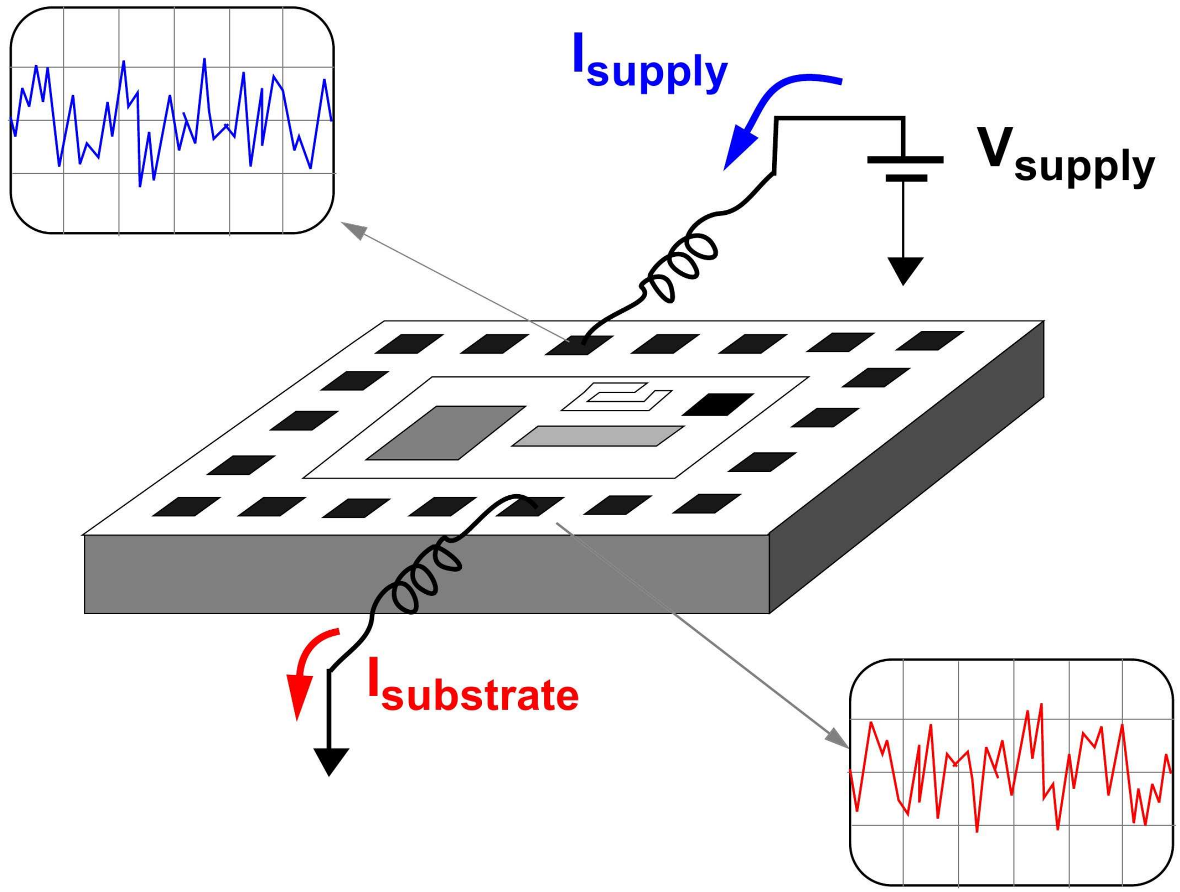
Units of phase noise





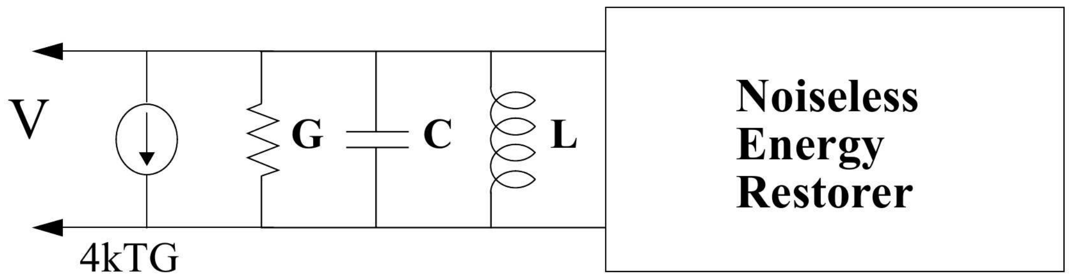
Phase noise units: dB below carrier, in a 1Hz bandwidth.

Substrate and supply noise



Phase noise: General considerations

- Idealize oscillator as *RLC* + negative resistance:



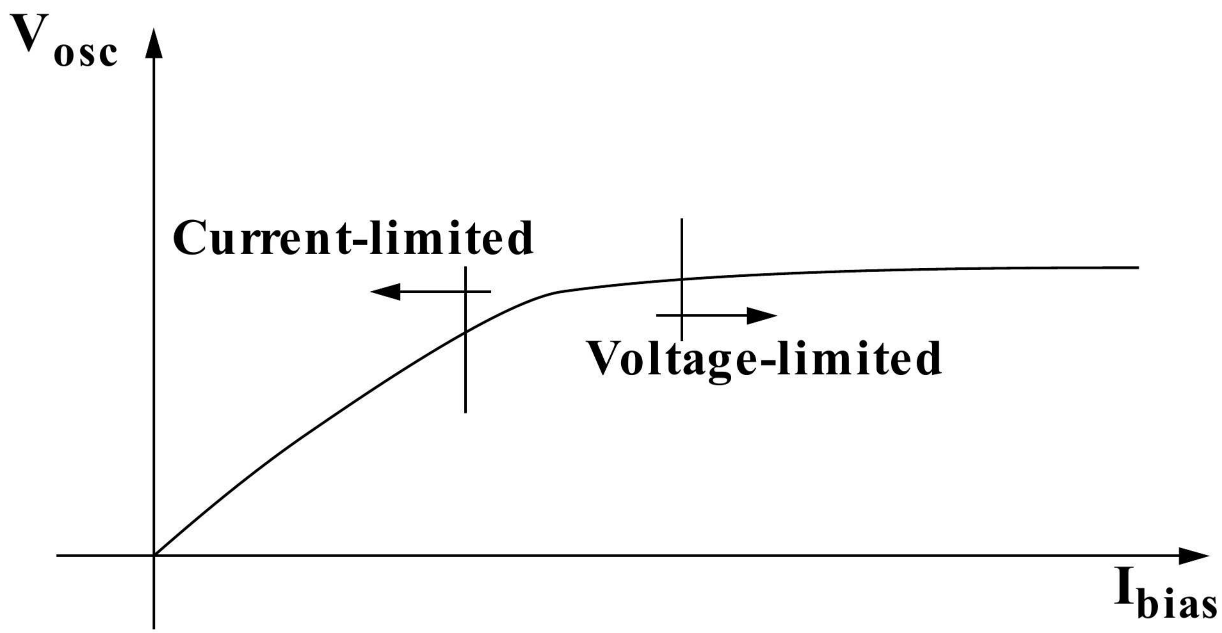
- Can show that the noise-to-signal ratio is

$$\frac{N}{S} = \frac{\overline{V_n^2}}{\overline{V_{sig}^2}} = \frac{kT}{E_{stored}} = \frac{\omega kT}{QP_{diss}}.$$

- Negative-*R* must cancel tank loss in steady state.
- Noise current then sees pure *LC* impedance.

Phase noise: General considerations

- Practical oscillators operate in one of two regimes:
  - *Current-limited*, where oscillation amplitude is proportional to $I_{bias}R_{tank}$.
  - *Voltage-limited*, where oscillation amplitude is independent of I_{bias}.



General considerations

- In the voltage-limited regime, increases in I_{bias} do not increase carrier power.
  - Additional bias current only increases noise and dissipation, so CNR degrades.
- In the current-limited regime, increases in I_{bias} increase signal power faster than noise power.
  - CNR increases until incipient voltage-limiting occurs.
- Best CNR occurs near boundary between limiting regimes.

Oscillator phase noise

- The expression for CNR reveals important optimization objectives:
  - $v_{carrier} \propto I_{bias}R_{tank} \Rightarrow P_{carrier} \propto (I_{bias})^2 R_{tank}$ in the current-limited regime.
  - $P_{noise} = kT/C = kT\omega^2 L$, if dominated by tank loss.
  - So $N/C \propto kT\omega^2 L/(I_{bias})^2 R_{tank}$ to an approximation.
- Generally want to *minimize* L/R to optimize for a given frequency and power consumption.
  - This contradicts much published advice to *maximize* L.
- N/C is important, but so is detailed spectrum.

Naive LTI model

- Assuming all noise comes from tank loss, PSD of tank voltage is approximately

$$\frac{\overline{v_n^2}}{\Delta f} = \frac{\overline{i_n^2}}{\Delta f}|Z|^2 = 4kTG\left(\frac{1}{G}\frac{\omega_0}{2Q\Delta\omega}\right)^2 = 4kTR\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2$$

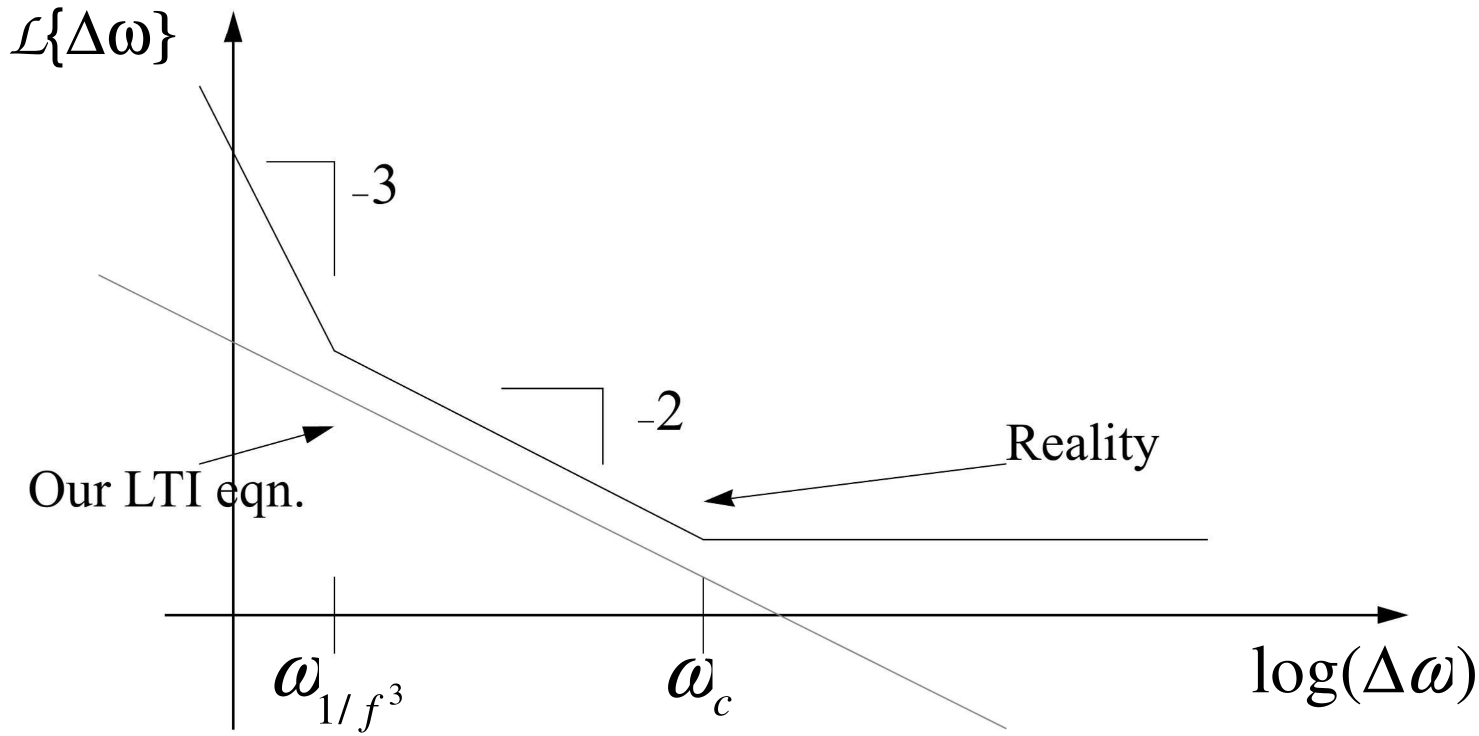
- Noise power splits evenly between phase and amplitude domains. Then

$$\mathcal{L}(\Delta\omega) = 10\log\left[\frac{\overline{v_n^2}/\Delta f}{\overline{v_{sig}^2}}\right] = 10\log\left[\frac{2kT}{P_{sig}}\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2\right]$$

- Funny units: dBc/Hz at a certain offset freq., e.g., "-110dBc/Hz@600kHz offset from 1.8GHz."

Naive LTI model vs. reality

- Previous expression doesn't describe real PN well:



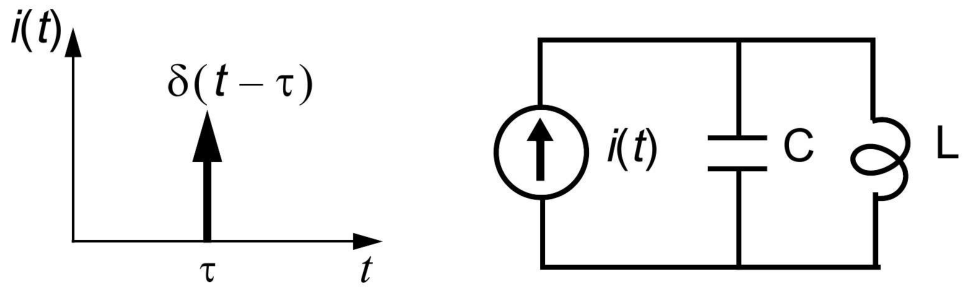
Leeson model

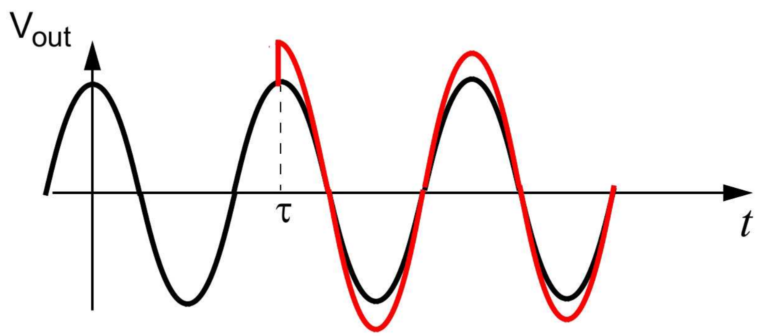
- Leeson provided quasi-empirical fix to remove discrepancies:

$$\mathcal{L}(\Delta\omega) = 10\log\left[\frac{2FkT}{P_{sig}}\left\{1+\left(\frac{\omega_0}{2Q\Delta\omega}\right)^2\right\}\left(1+\frac{\Delta\omega_{1/f^3}}{|\Delta\omega|}\right)\right]$$

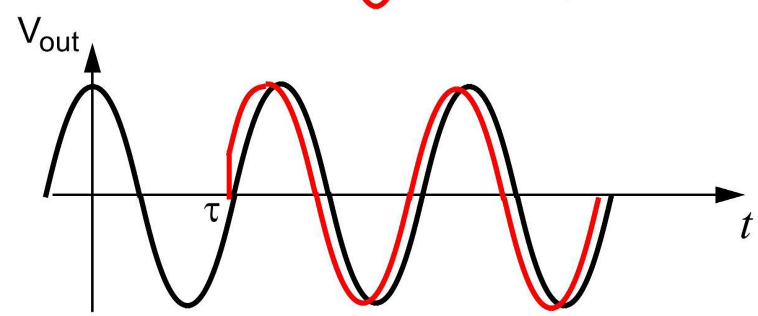
- *F* accounts for excess noise at all offsets.
- $\Delta\omega_{1/f^3}$ accounts for $1/f^3$ region near carrier.
- First additive factor of 1 accounts for noise floor.
- These factors cannot be determined *a priori*; they are *a posteriori* fitting parameters.
- Need to revisit assumption that oscillators are LTI.

Are oscillators LTI?





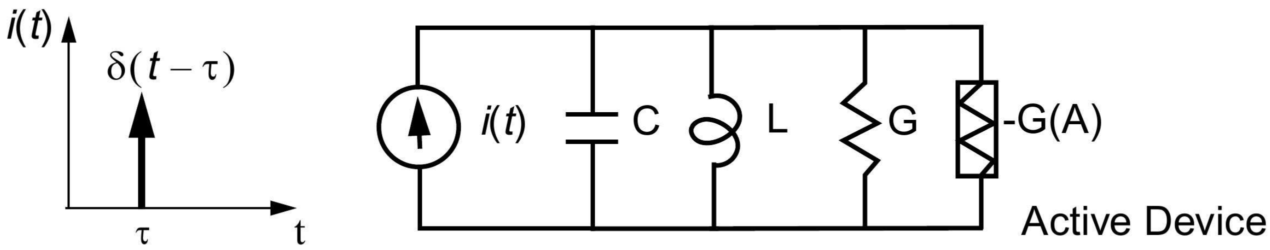
Impulse injected at peak

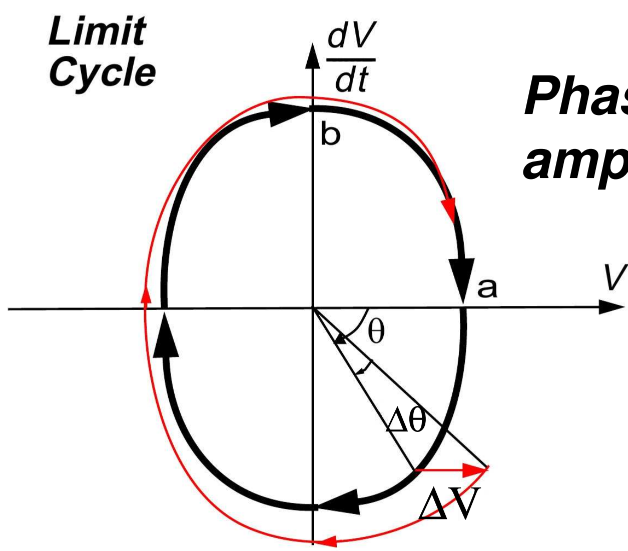


Impulse injected at zero crossing

Even for an ideal LC, the phase response is time-variant.

Amplitude restoration in real oscillators

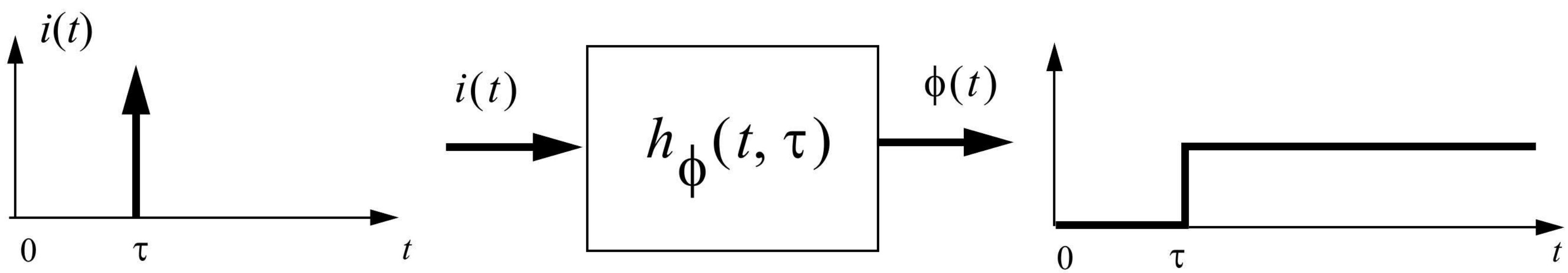




Phase errors persist indefinitely; amplitude errors quench over time

Phase impulse response

The phase impulse response of an oscillator is a step:



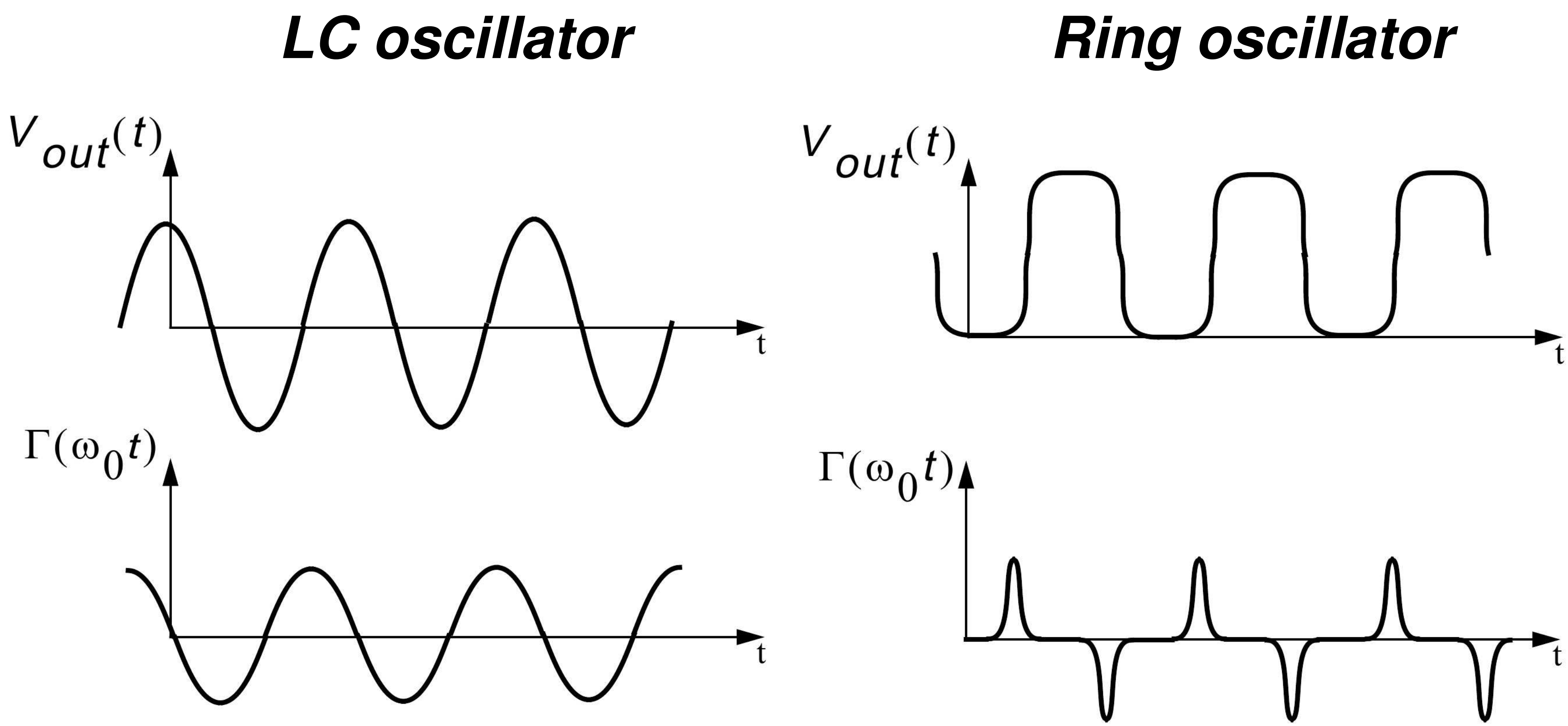
The unit impulse response is:

$$h_\phi(t,\tau) = \frac{\Gamma(\omega_0 t)}{q_{\max}} u(t-\tau)$$

Γ(x) is a dimensionless function, periodic in 2π, describing how much phase change results from impulse at

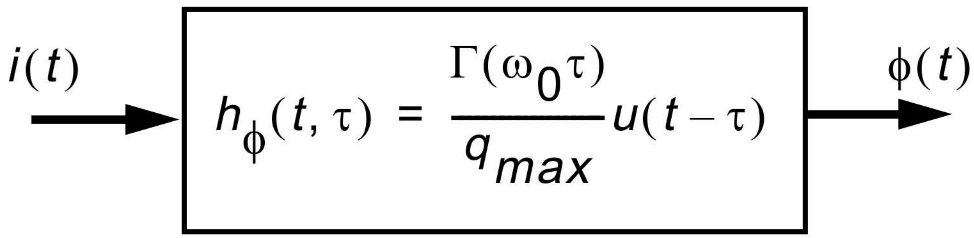
$$t = T\frac{x}{2\pi}$$

Impulse sensitivity function (ISF)



The ISF quantifies the sensitivity to perturbations at all instants.

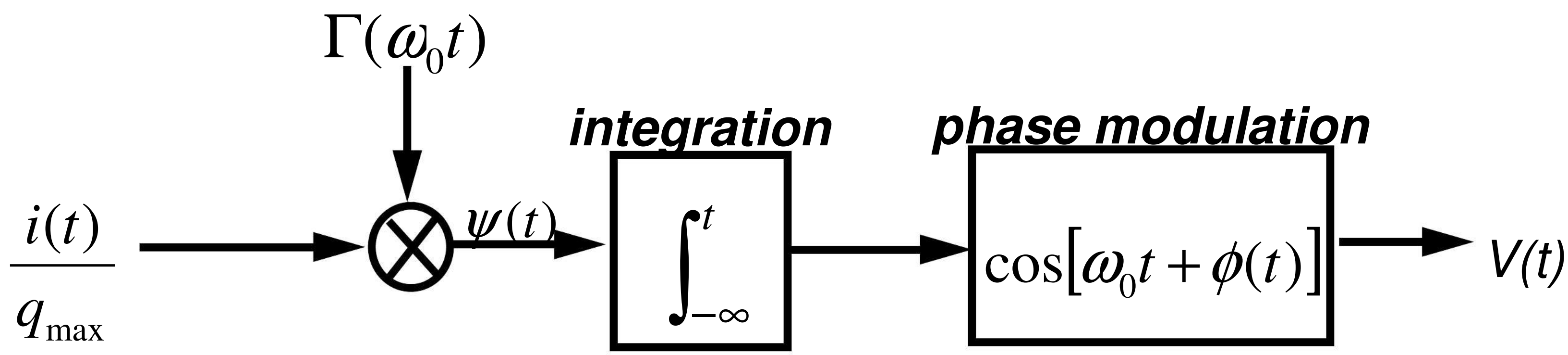
Phase response to arbitrary inputs



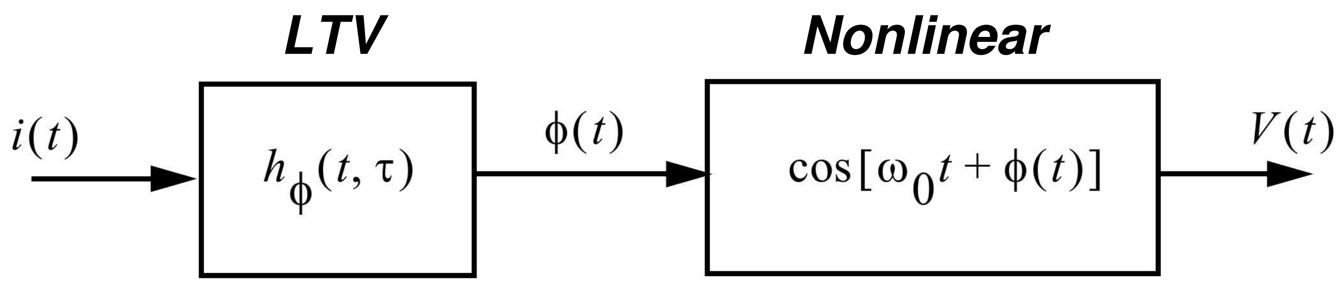
Use the superposition integral to compute phase response:

$$\phi(t) = \int_{-\infty}^{\infty} h_\phi(t,\tau)i(\tau)d\tau = \frac{1}{q_{\max}}\int_{-\infty}^{t}\Gamma(\omega_0 t)i(\tau)d\tau$$

Block diagram representation of superposition integral is:



Phase noise due to white noise



For a white noise input current of spectral density $\dfrac{\overline{i_n^2}}{\Delta f}$,
the phase noise is given by

$$\mathcal{L}(\Delta\omega) = 10\log\left[\frac{\Gamma_{rms}^2}{q_{\max}^2}\,\frac{\overline{i_n^2}/\Delta f}{2(\Delta\omega)^2}\right],$$

where Γ_{rms} ***is the rms value of the ISF.***

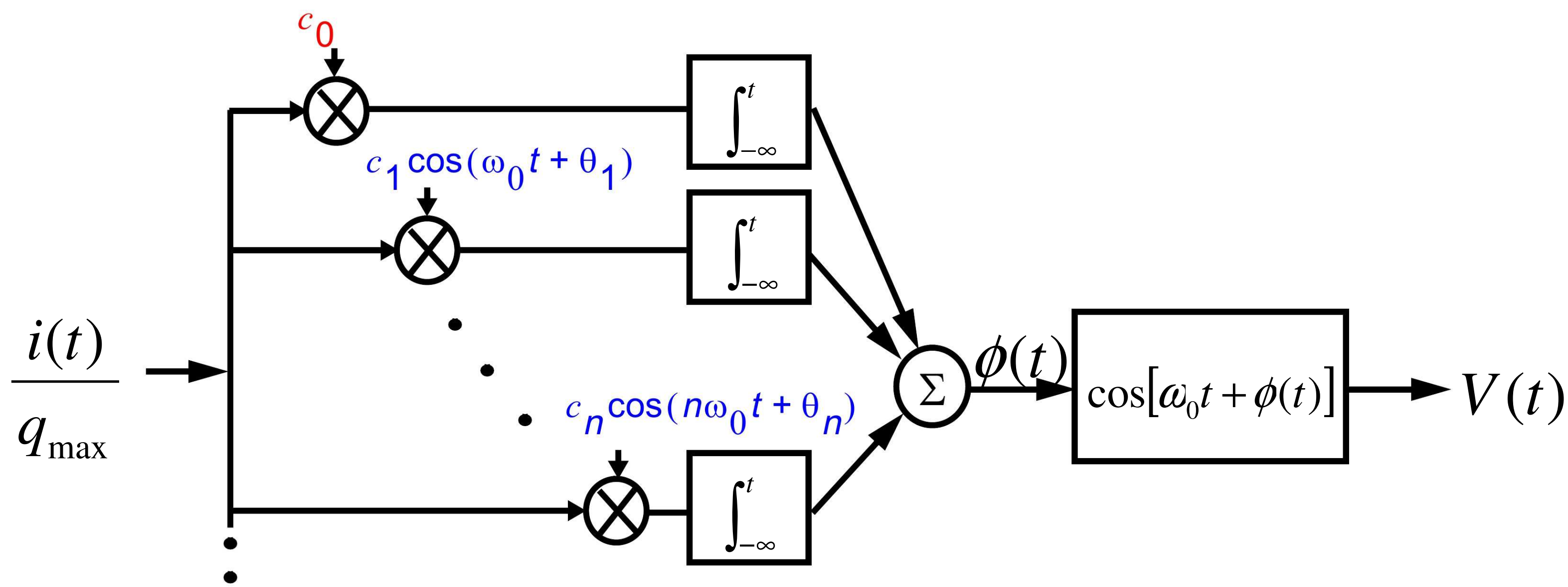
ISF in greater detail

ISF is periodic, expressible as a Fourier series:

$$\Gamma(\omega_0 t) = c_0 + \sum_{1}^{\infty} c_n \cos(n\omega_0 t + \theta_n)$$

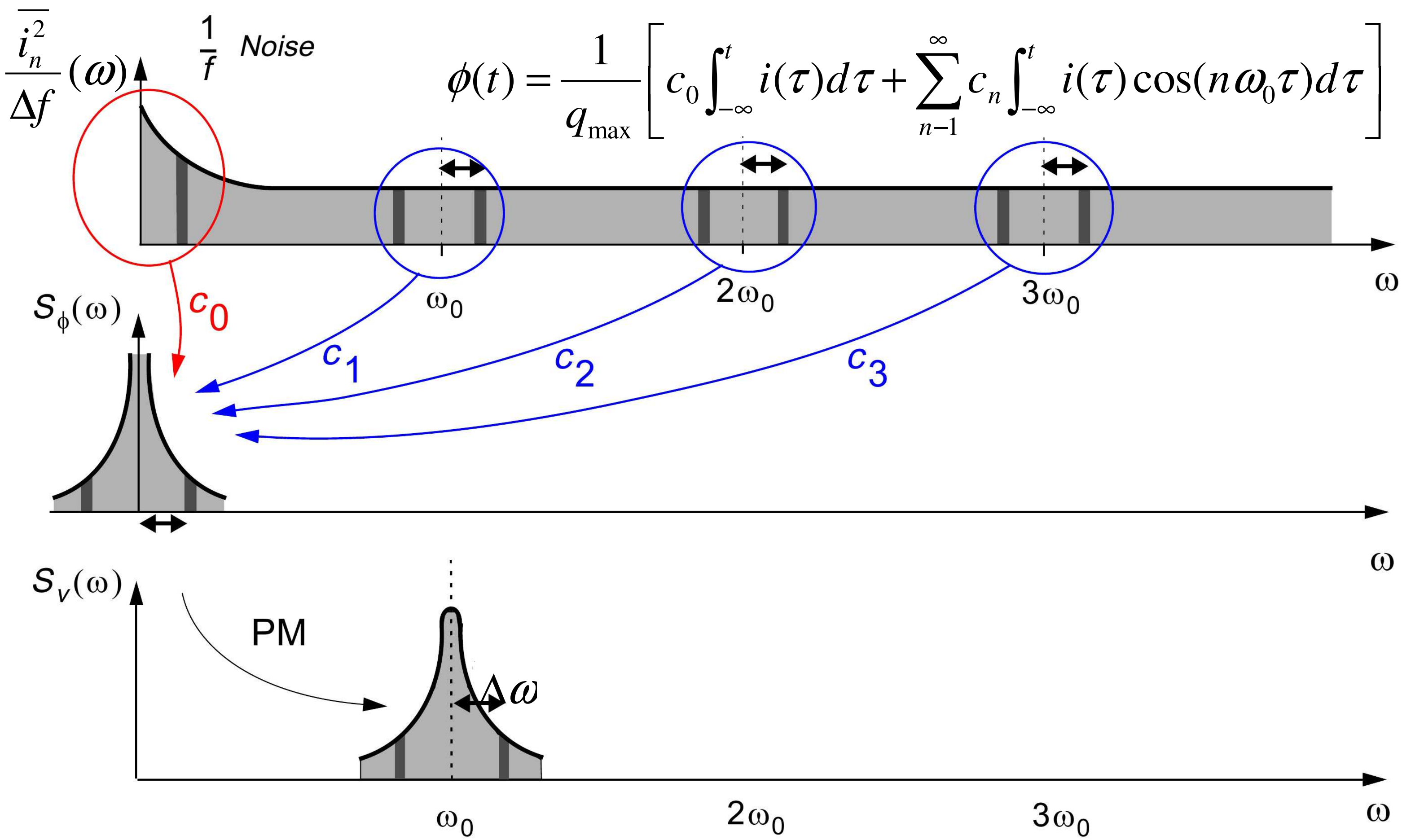
The phase is then as follows:

$$\phi(t) = \frac{1}{q_{\max}}\left[c_0 \int_{-\infty}^{t} i(\tau)\,d\tau + \sum_{n-1}^{\infty} c_n \int_{-\infty}^{t} i(\tau)\cos(n\omega_0 \tau)\,d\tau \right]$$



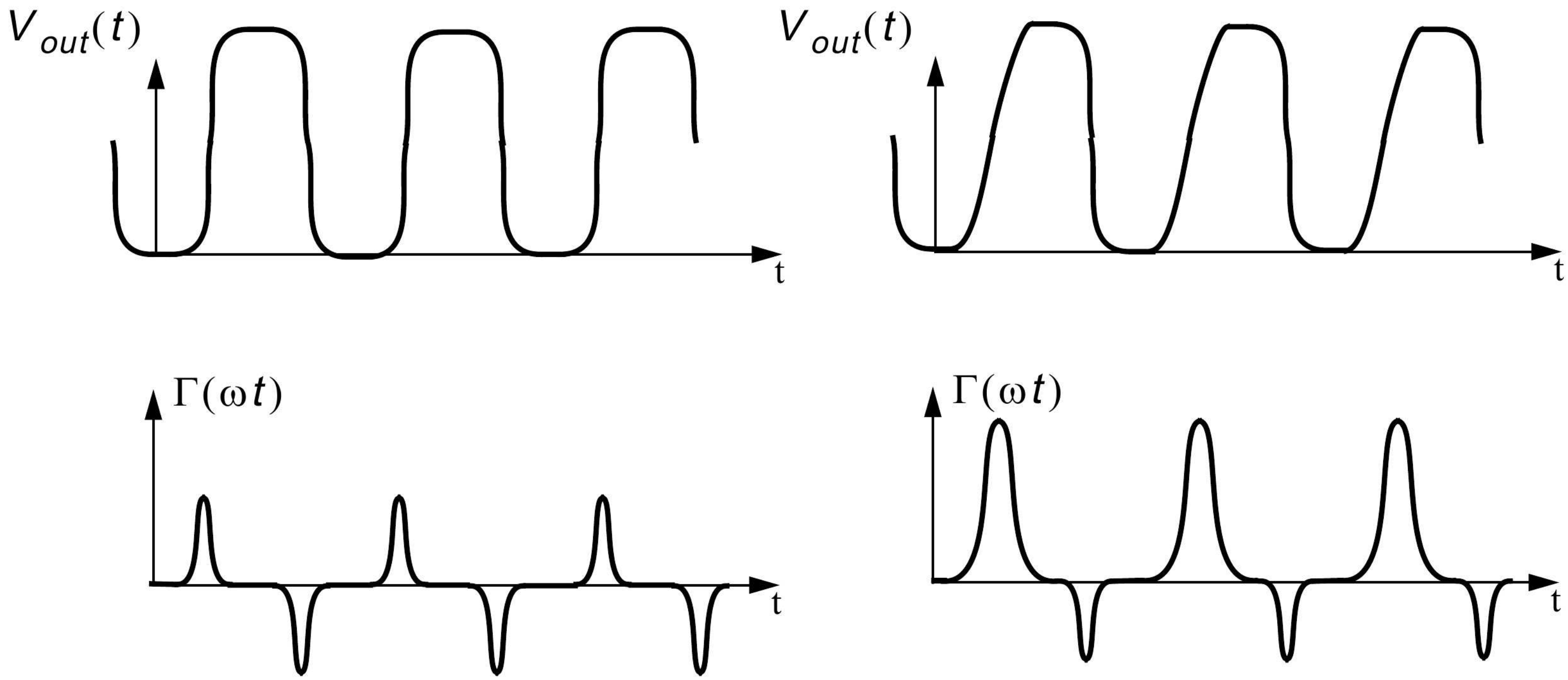
Contributions by noise at $n\omega_0$

$$\phi(t) = \frac{1}{q_{\max}}\left[c_0\int_{-\infty}^{t} i(\tau)\,d\tau + \sum_{n-1}^{\infty} c_n \int_{-\infty}^{t} i(\tau)\cos(n\omega_0\tau)\,d\tau\right]$$



Symmetry matters

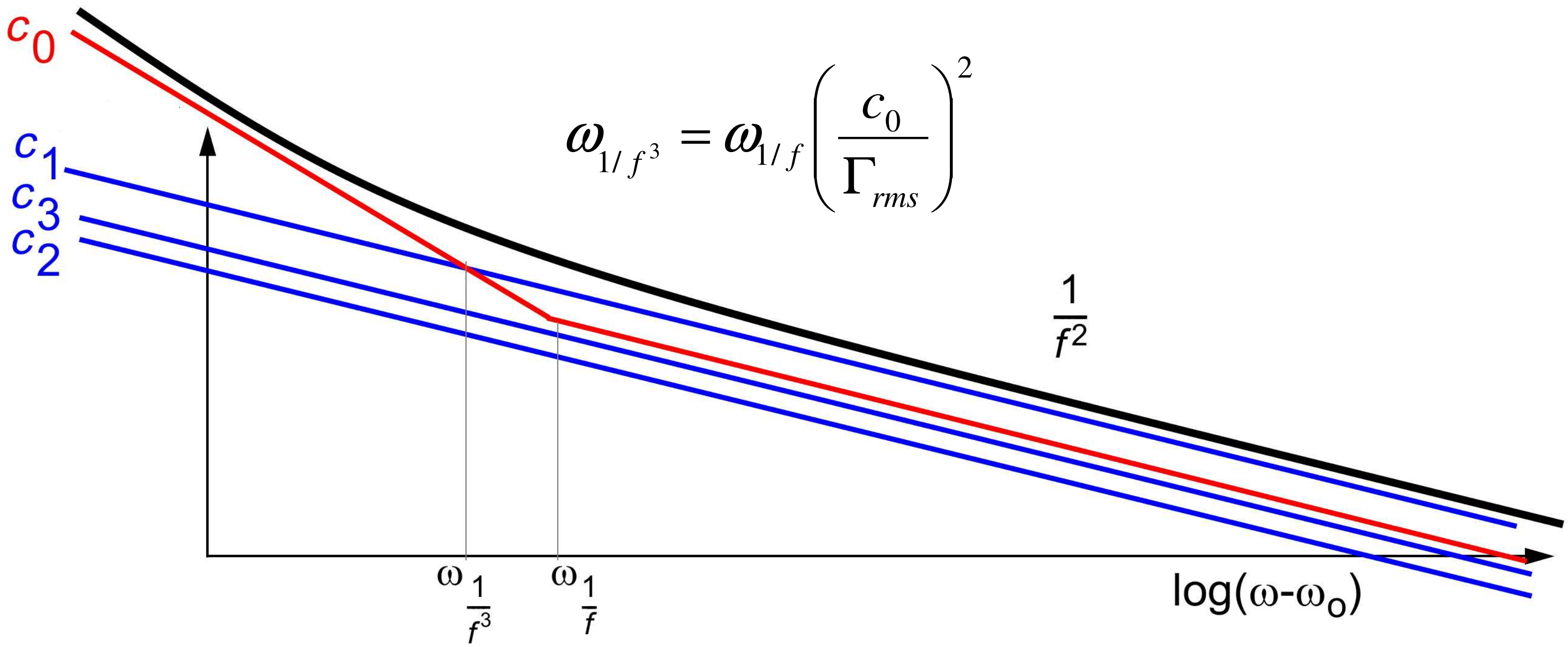
$$c_0 = \frac{1}{2\pi}\int_0^{2\pi}\Gamma(x)dx$$



The dc value of the ISF is affected by waveform symmetry

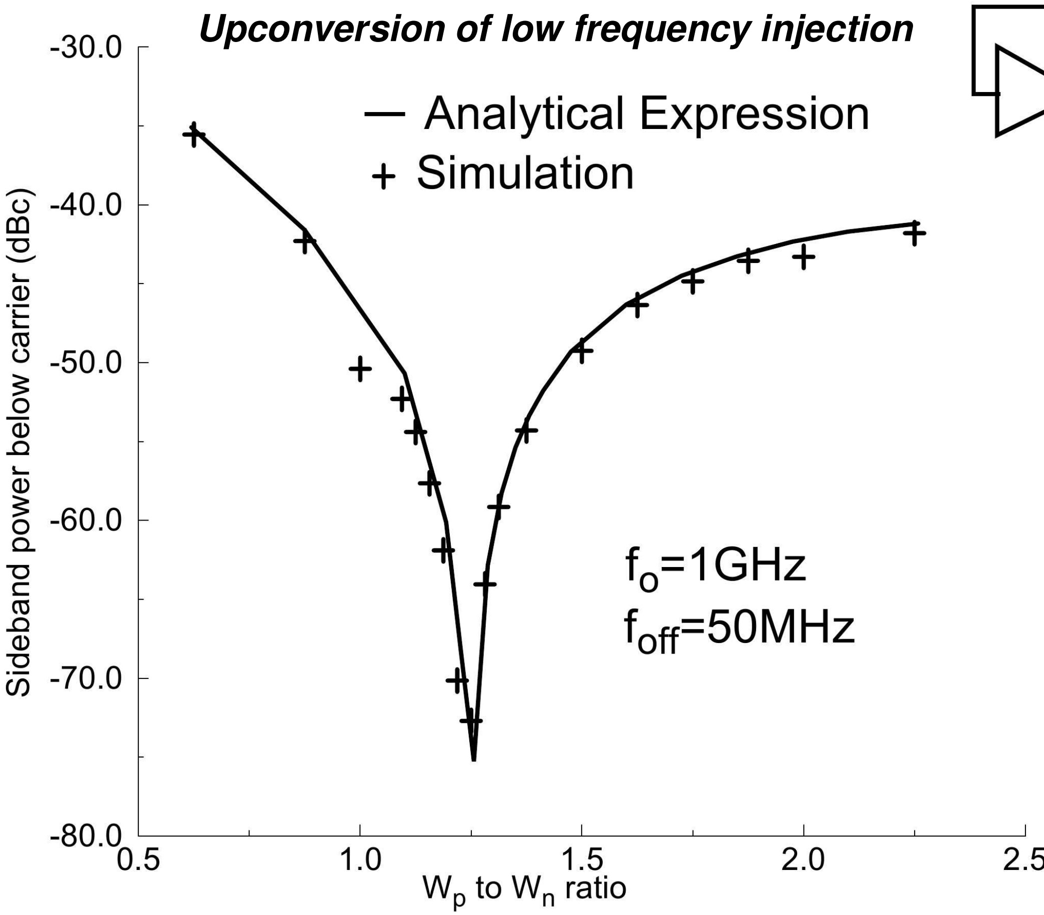
1/f^3 corner of phase noise spectrum

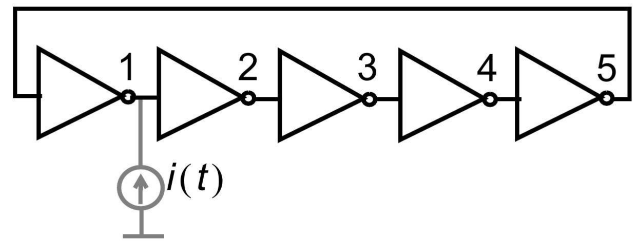
The 1/f³ phase noise corner is not ***the same as the 1/f device noise corner.***

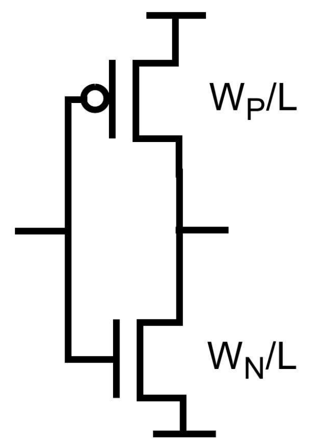


Adjustment of waveform symmetry can control 1/f noise upconversion.

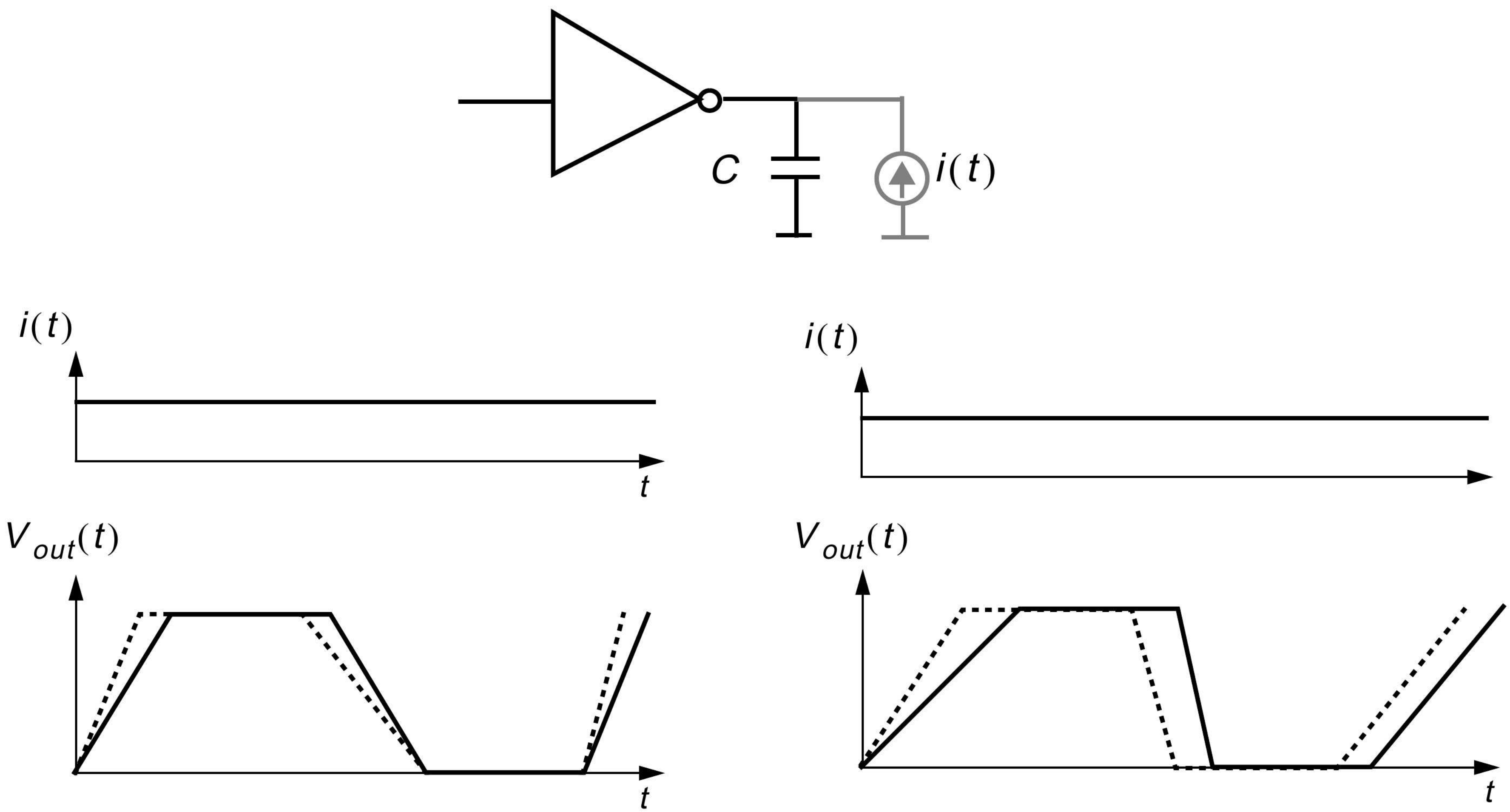
Exploring the effect of waveform symmetry





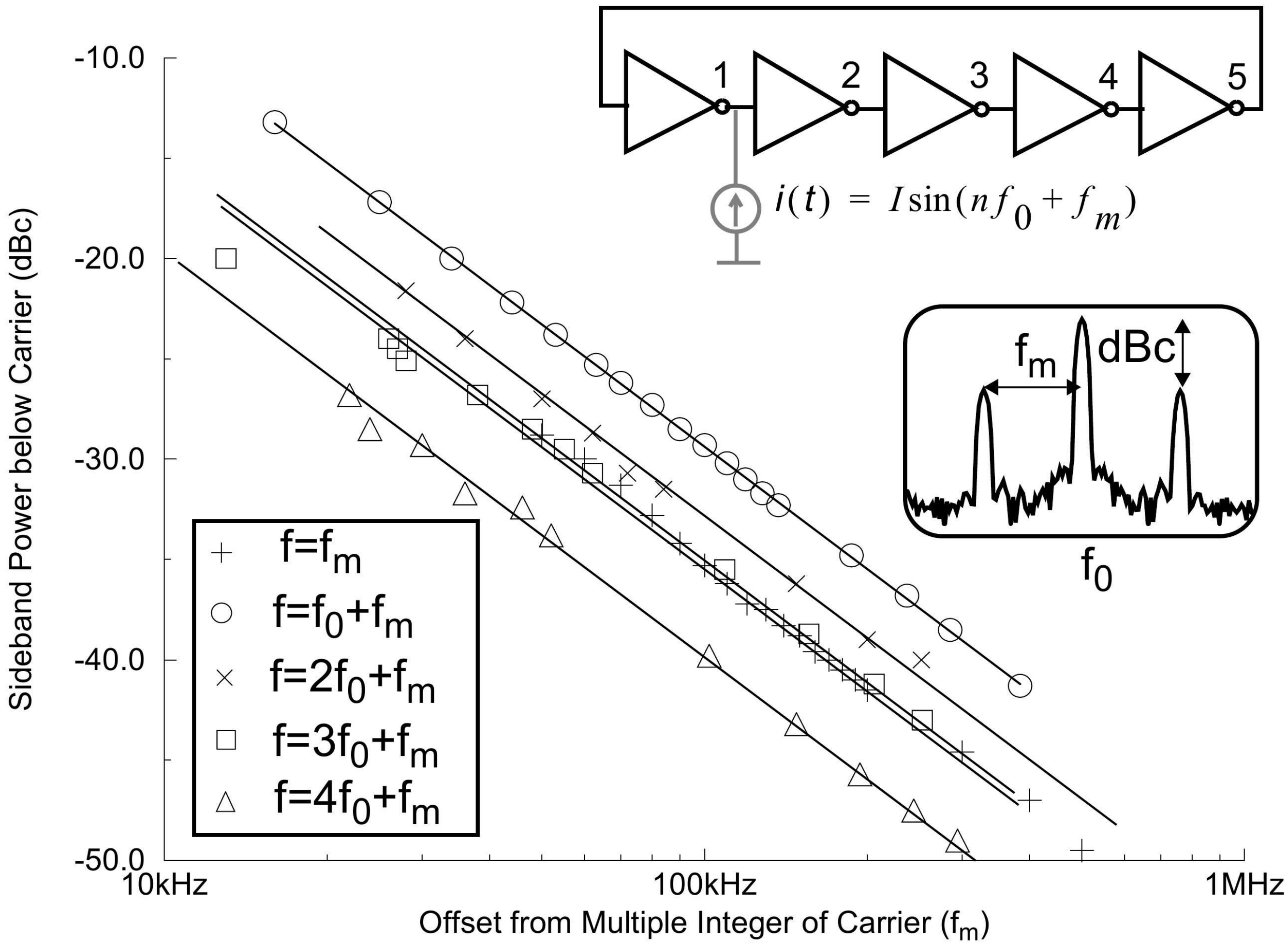


Effect of waveform symmetry

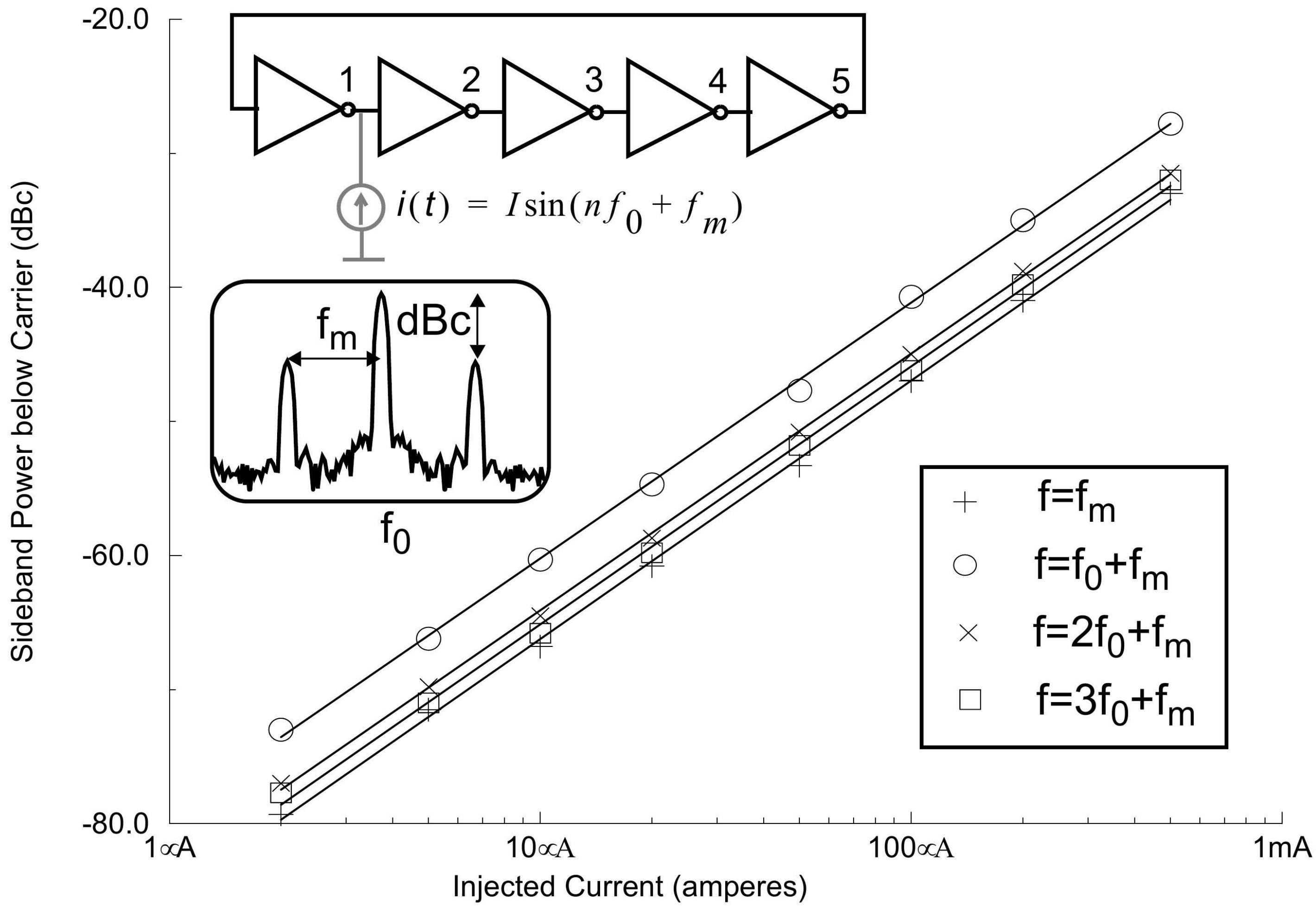


Low frequency injection shifts the frequency of the asymmetric waveform.

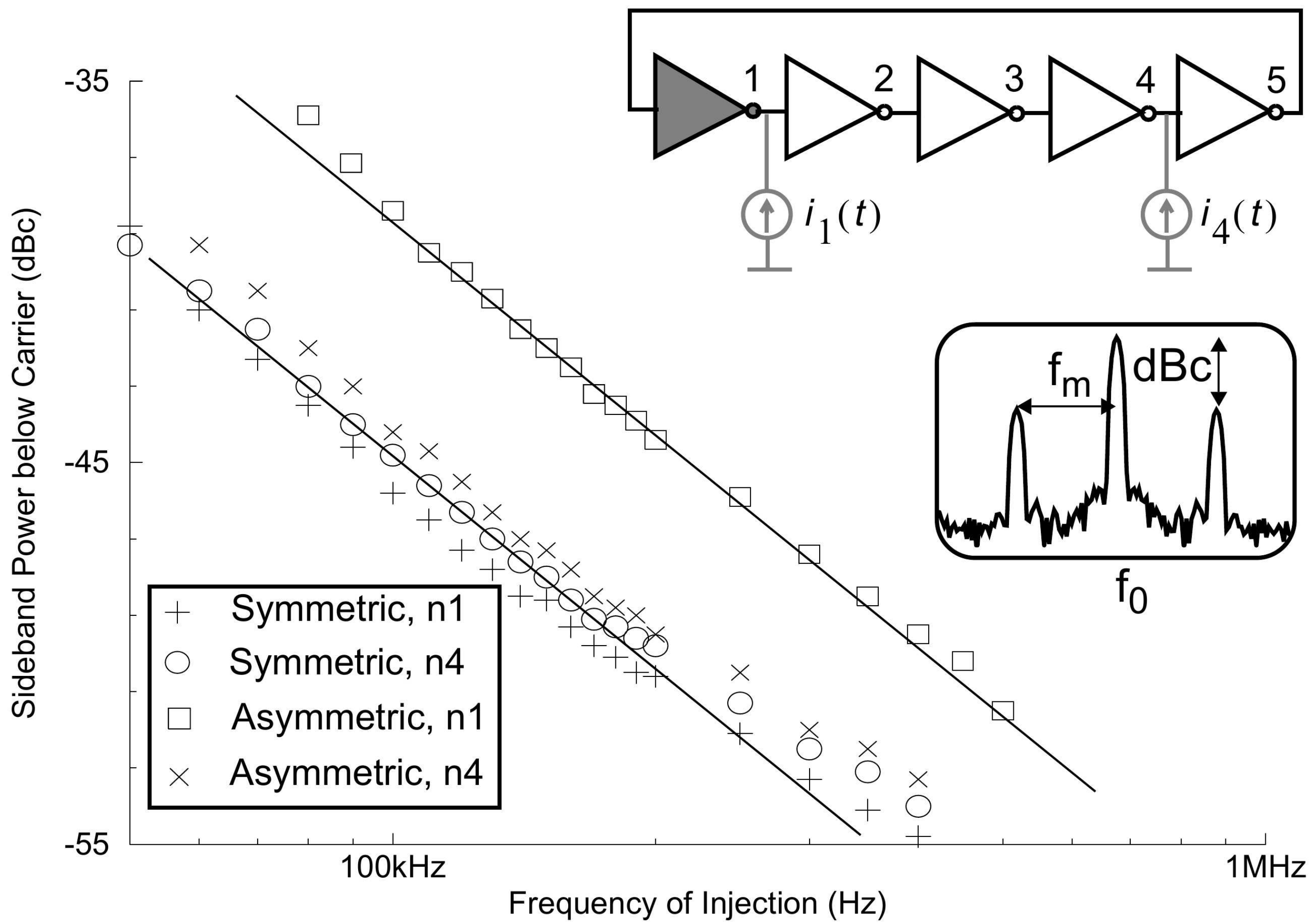
Effect of injection at nf_0



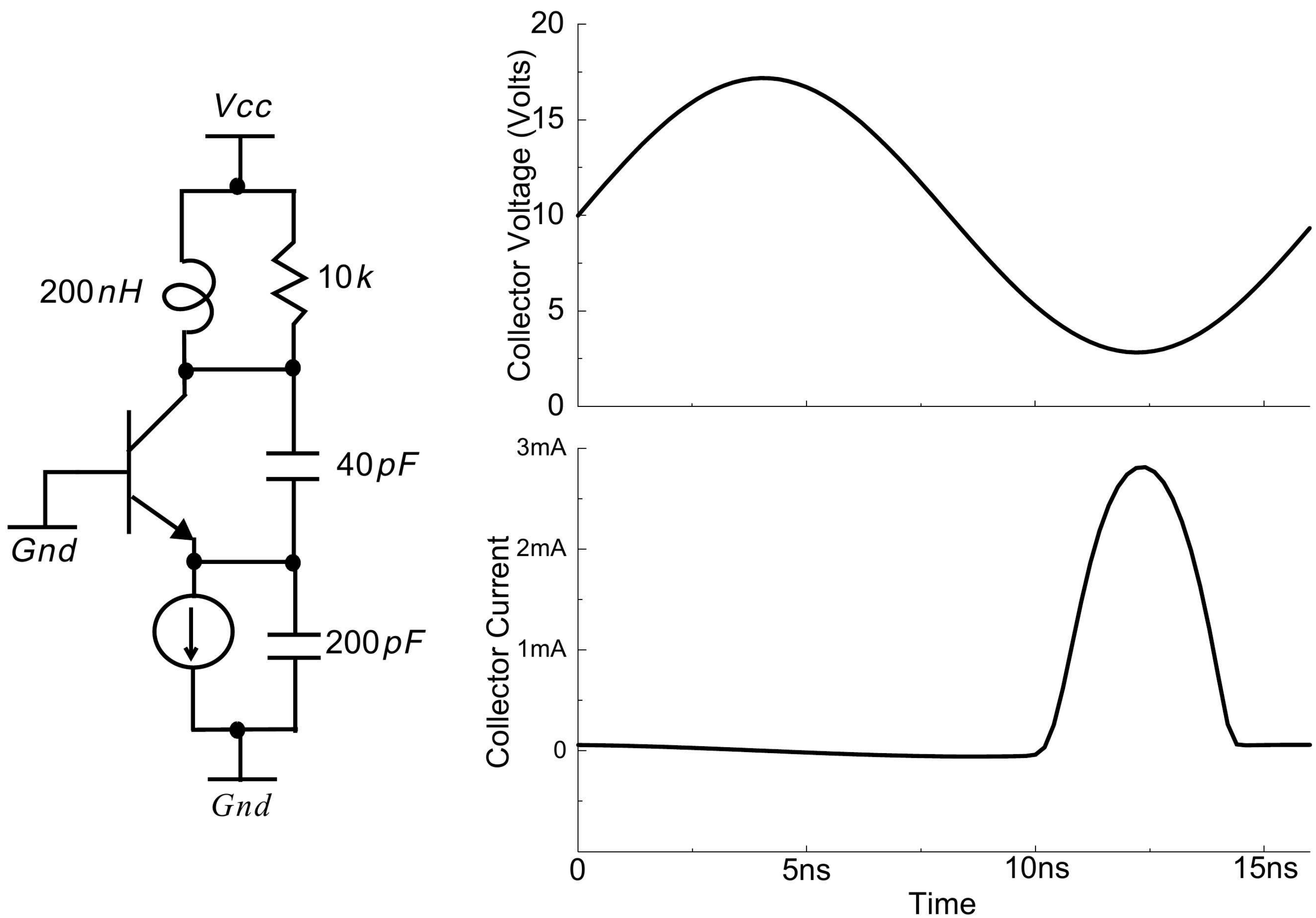
Sideband power vs. injected current



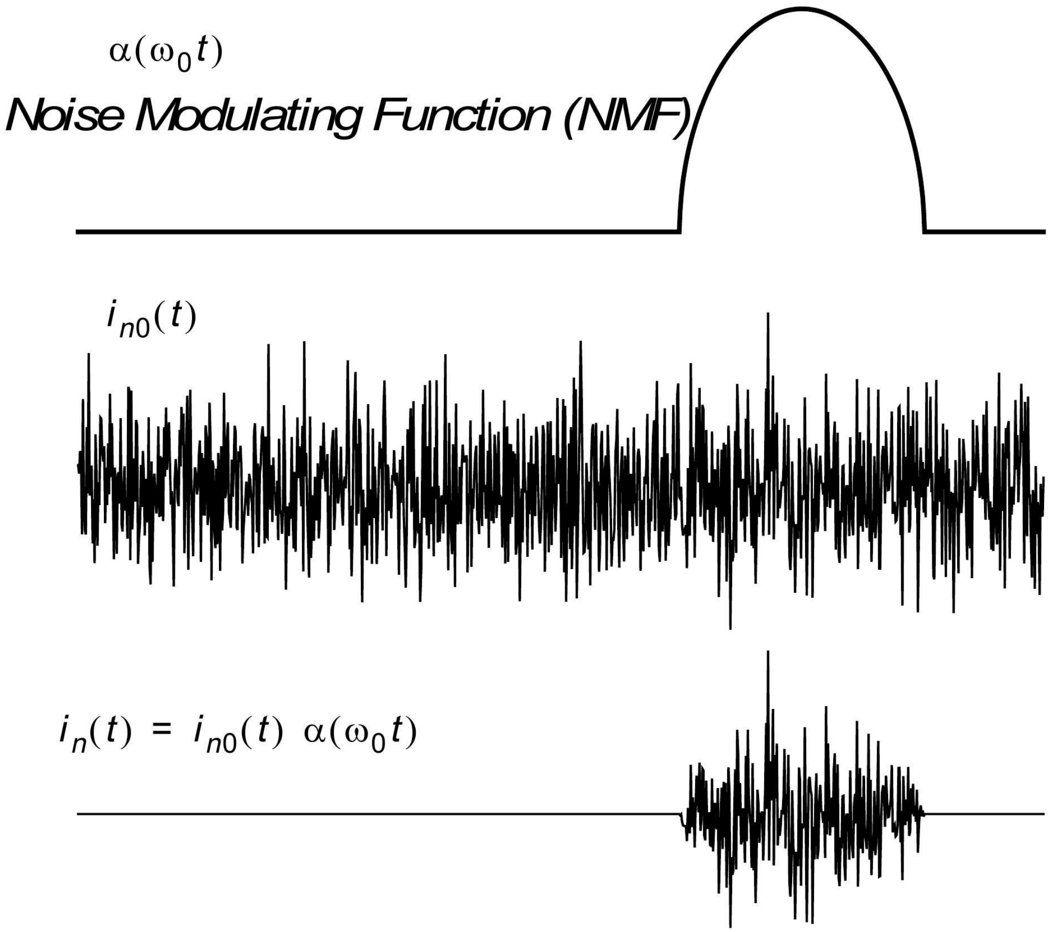
Symmetric vs. asymmetric ring oscillator



Oscillator currents are time-varying



Effects of time-varying currents



$$\phi(t) = \int_{-\infty}^{t} i_n(t)\frac{\Gamma(\omega_0\tau)}{q_{\max}}d\tau$$

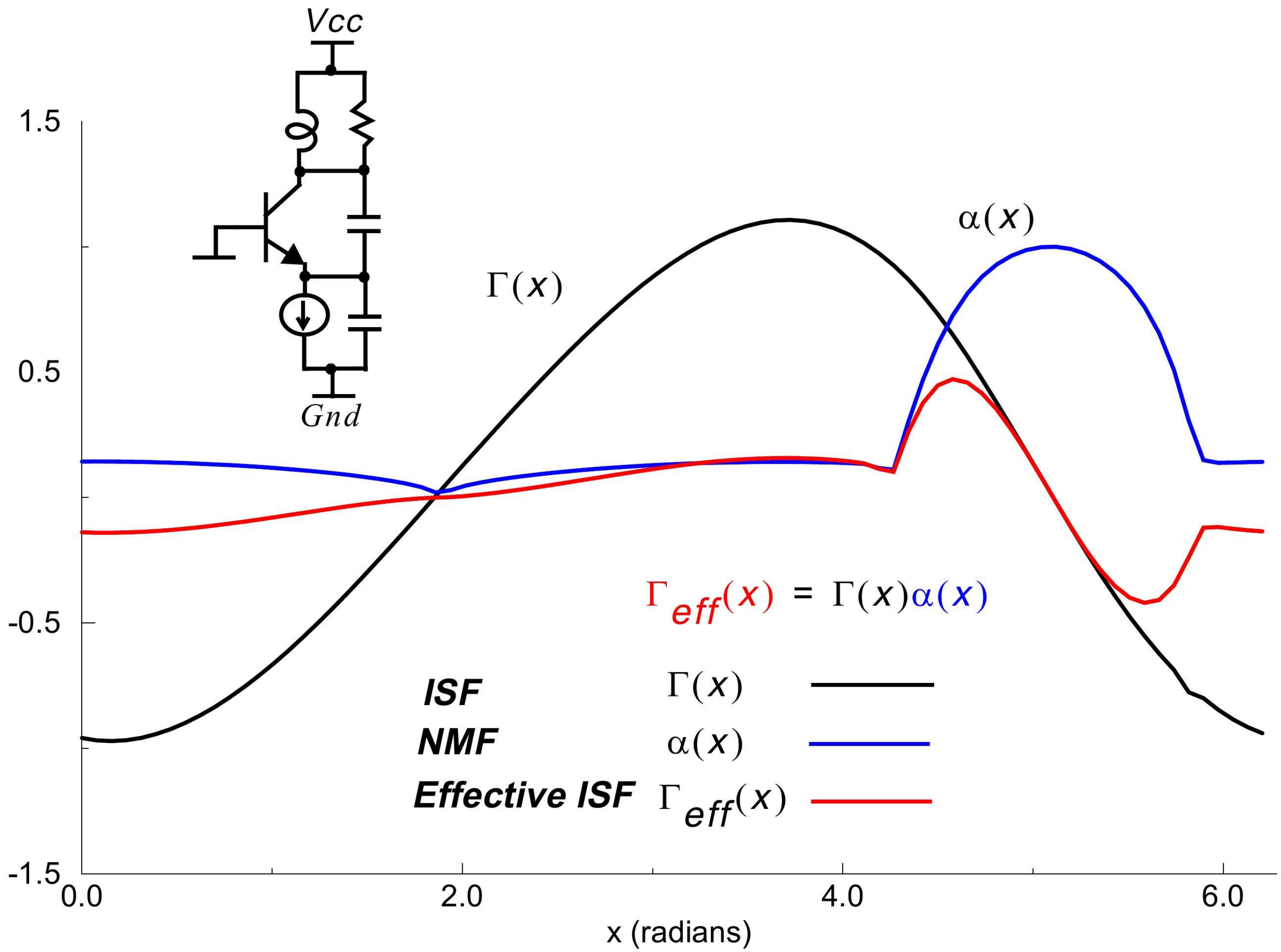
$$= \int_{-\infty}^{t} i_{n0}(t)\frac{\alpha(\omega_0\tau)\Gamma(\omega_0\tau)}{q_{\max}}d\tau$$

Effective ISF:

$$\Gamma_{eff}(x) = \Gamma(x)\cdot\alpha(x)$$

Cyclostationary noise can be modeled as stationary with modified ISF.

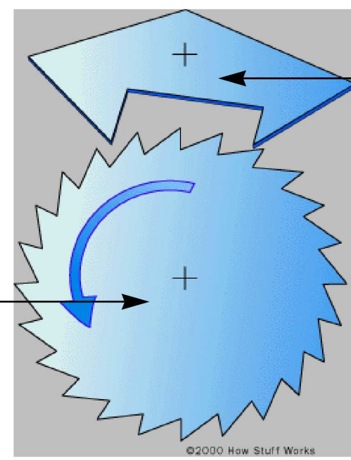
Colpitts



Vcc
Gnd
1.5
0.5
-0.5
-1.5
0.0
2.0
4.0
6.0
x (radians)
Γ(x)
α(x)
Γ_eff(x) = Γ(x)α(x)
ISF Γ(x)
NMF α(x)
Effective ISF Γ_eff(x)

Plus ça change...

- To exploit cyclostationarity, deliver energy to the tank impulsively, where the ISF is a minimum.
- This idea is old; an *escapement* in mechanical clocks delivers energy to pendulum impulsively.



coupled to pendulum

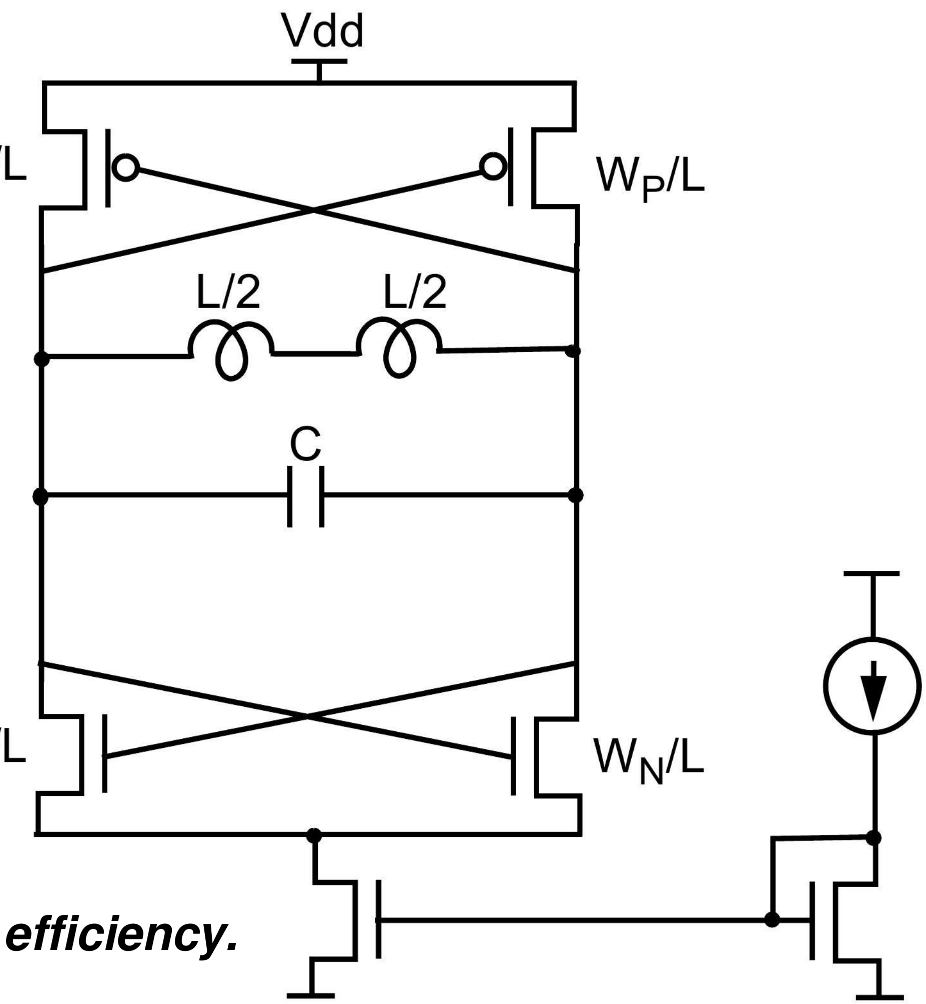
powered by spring

Conditions for optimal phase noise

- Impulses are delivered at or near pendulum's velocity maxima.
  - The escapement returns energy with minimal perturbation of oscillation period.[Airy, 1826]

- Similarly, the optimal moments in an *LC* oscillator are near the tank voltage maxima.

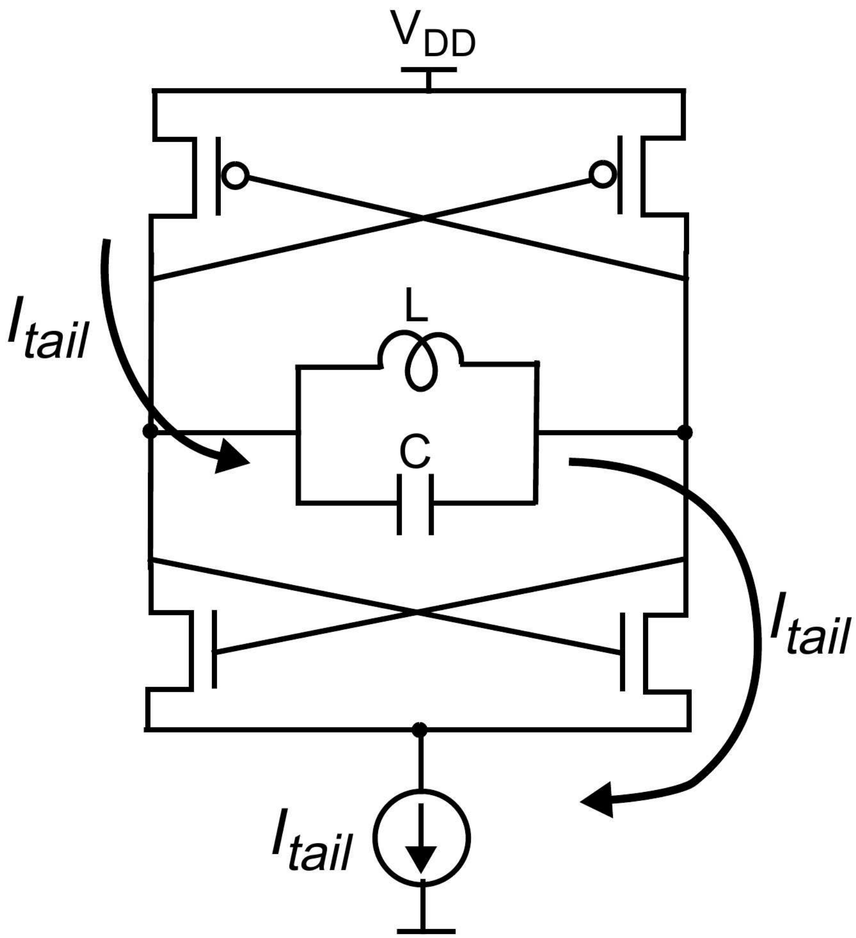
A symmetric *LC* oscillator



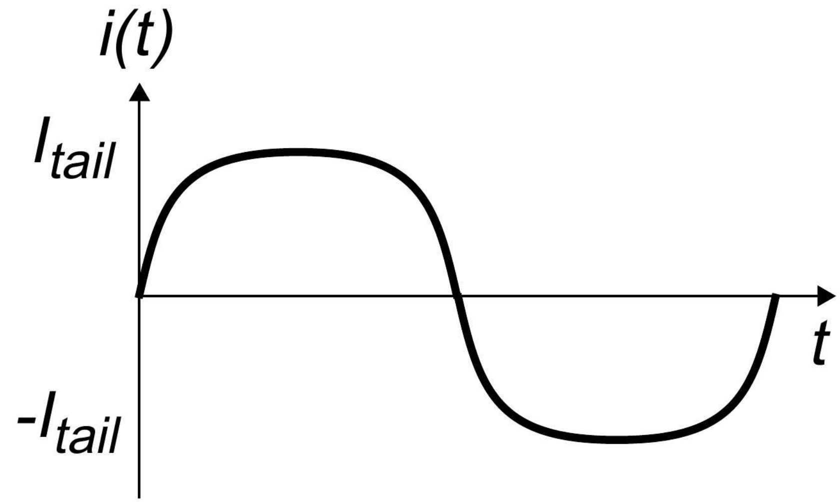
Adjust width ratio to minimize dc value of ISF

"Reuse" of current improves efficiency. [Craninckx et al., CICC 1997]

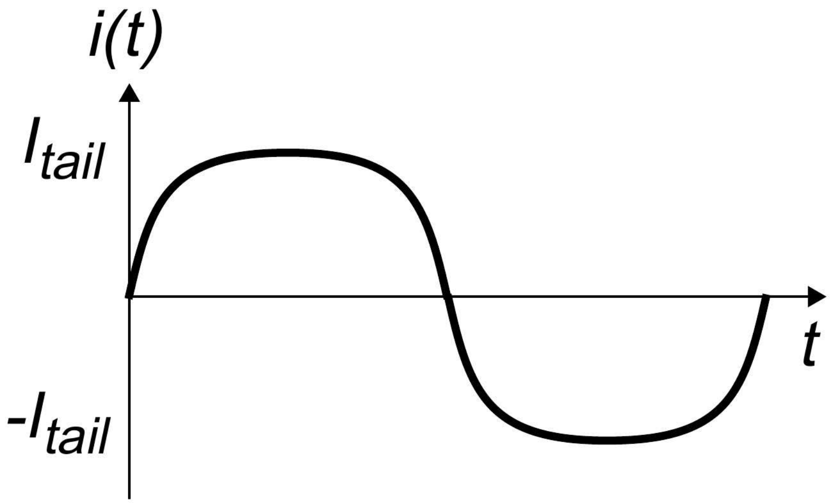
Tank voltage amplitude

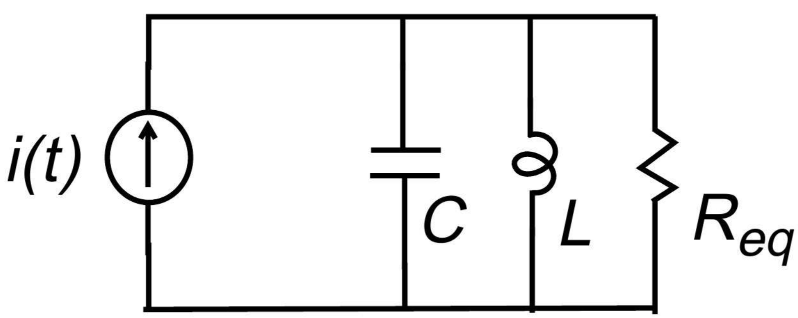


Assuming fast current steering, current can be approximated as nearly square:



Tank voltage amplitude





For an ideal square waveform:

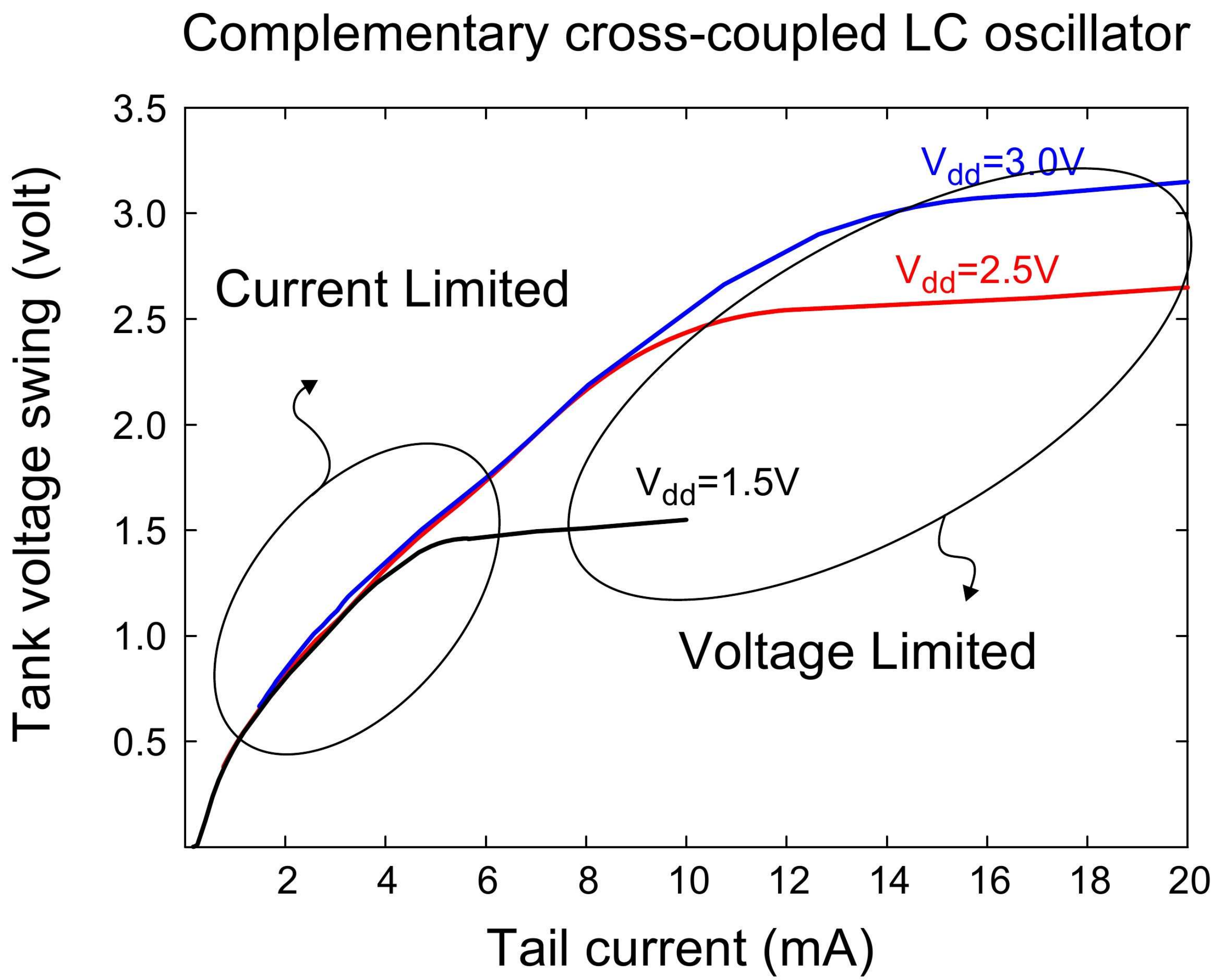
$$V_{\max} = \frac{4}{\pi} I_{tail} R_{eq}$$

For a sinusoidal waveform:

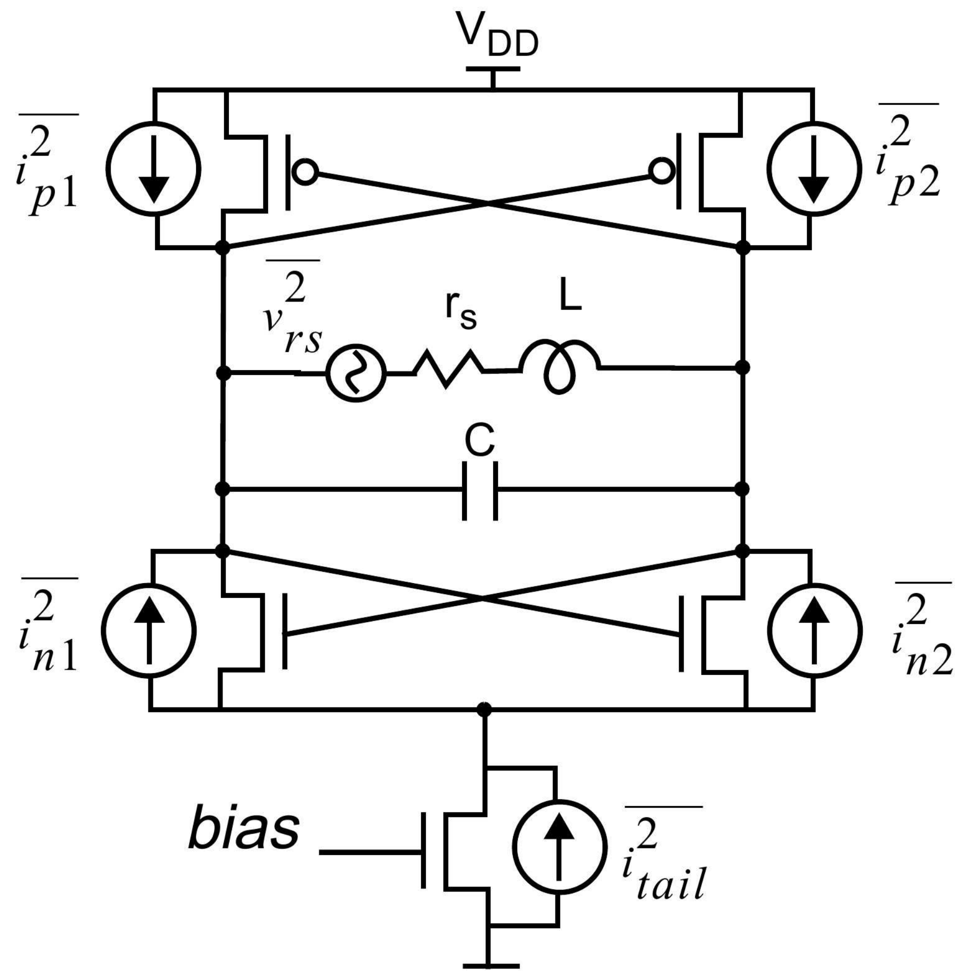
$$V_{\max} = I_{tail} R_{eq}$$

(assuming operation in current-limited mode)

Modes of amplitude limiting



Major noise sources



Not all noise sources affect phase noise equally.

MOS noise:

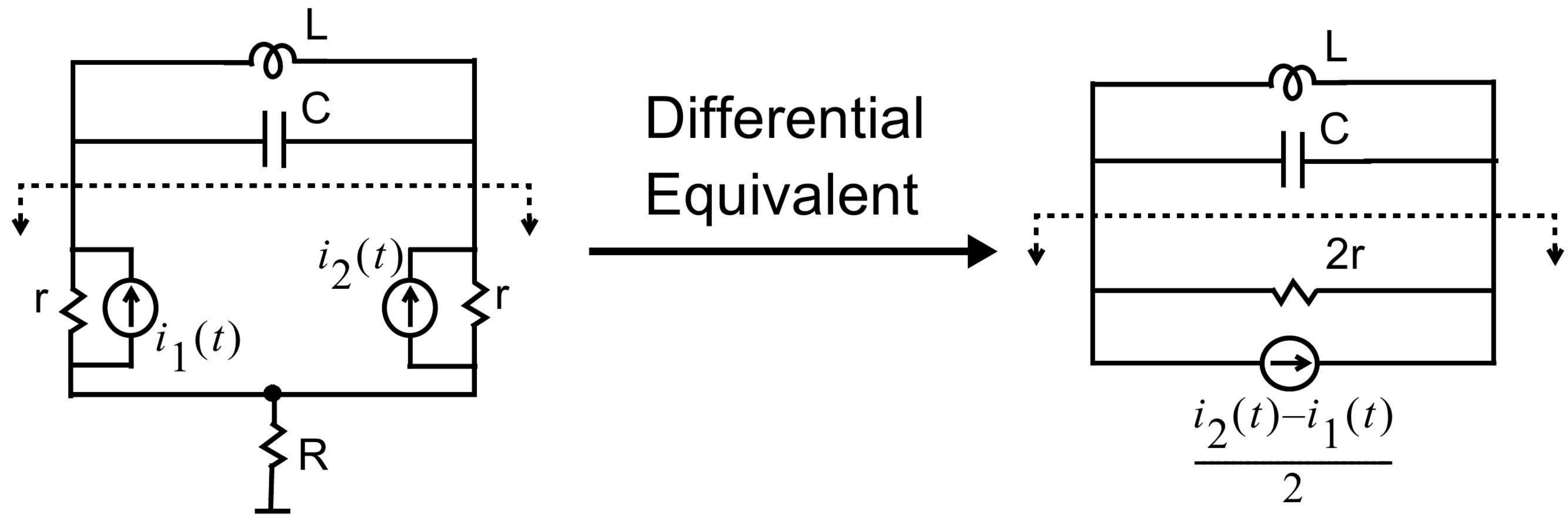
$$\frac{\overline{i_n^2}}{\Delta f} = 4kT\gamma\mu C_{ox}\frac{W}{L}\left(V_{GS} - V_T\right),$$

(valid for long- and short-channel devices).

Inductor noise:

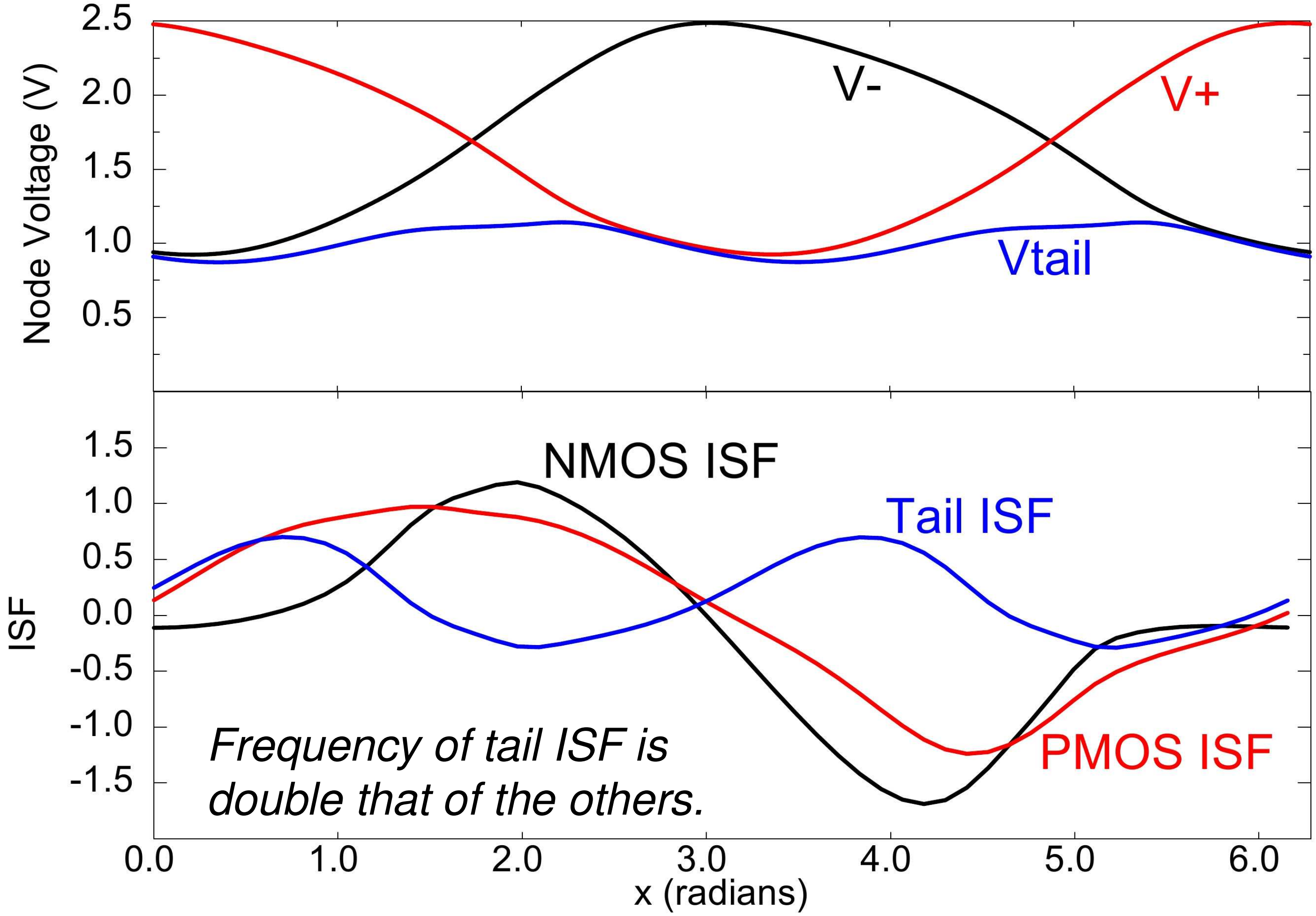
$$\frac{\overline{v_n^2}}{\Delta f} = 4kTr_s.$$

Equivalent circuit for noise sources

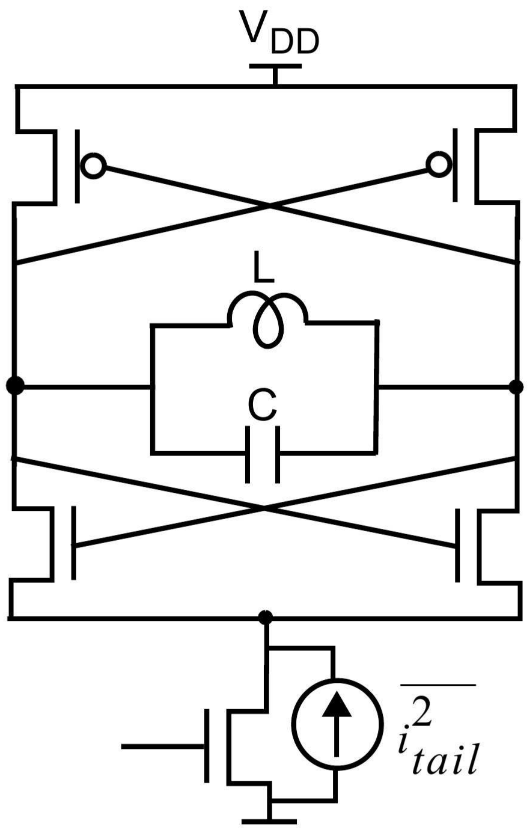


$$\left.\frac{\overline{i_n^2}}{\Delta f}\right|_{diff-pair} = \frac{1}{4}\left(\frac{\overline{i_{n1}^2}}{\Delta f} + \frac{\overline{i_{n2}^2}}{\Delta f} + \frac{\overline{i_{p1}^2}}{\Delta f} + \frac{\overline{i_{p2}^2}}{\Delta f}\right) = \frac{1}{2}\left(\frac{\overline{i_n^2}}{\Delta f} + \frac{\overline{i_p^2}}{\Delta f}\right)$$

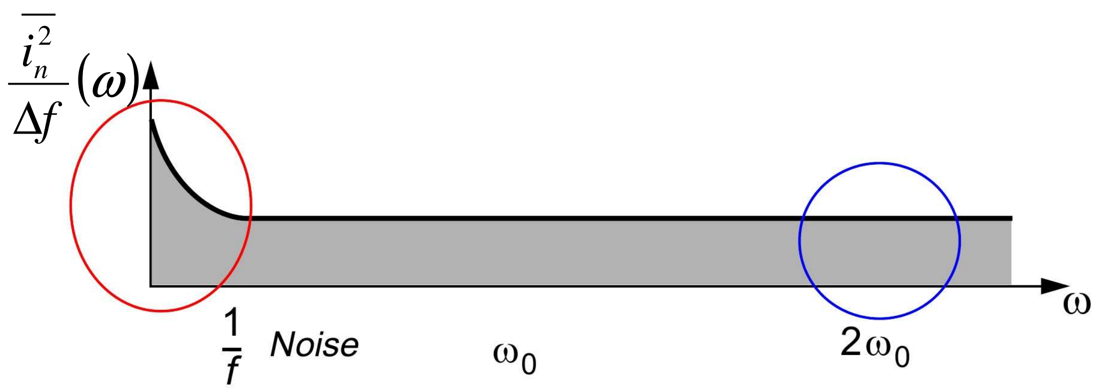
Waveform and ISF



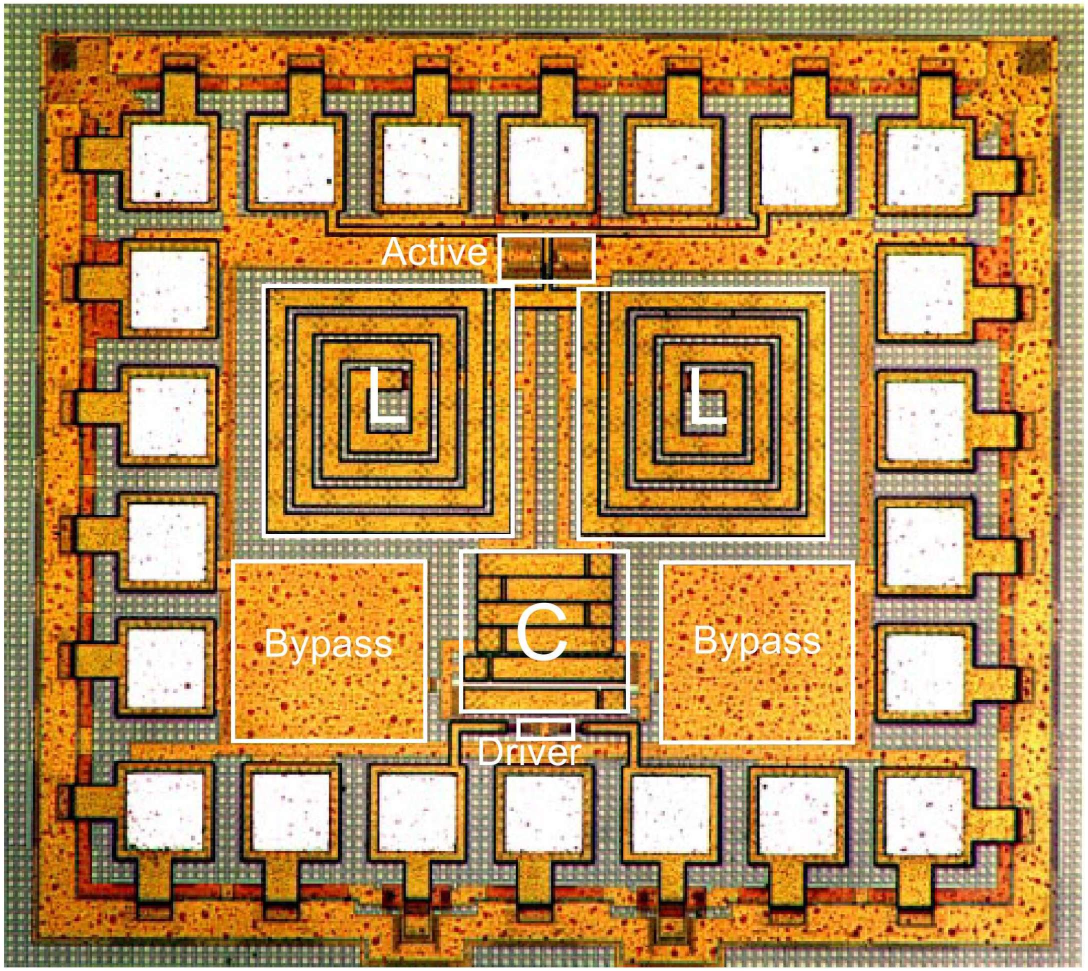
Effect of tail current noise



For the tail current source, only noise centered about even harmonics of the carrier contribute to phase noise, as consequence of double-frequency tail ISF.

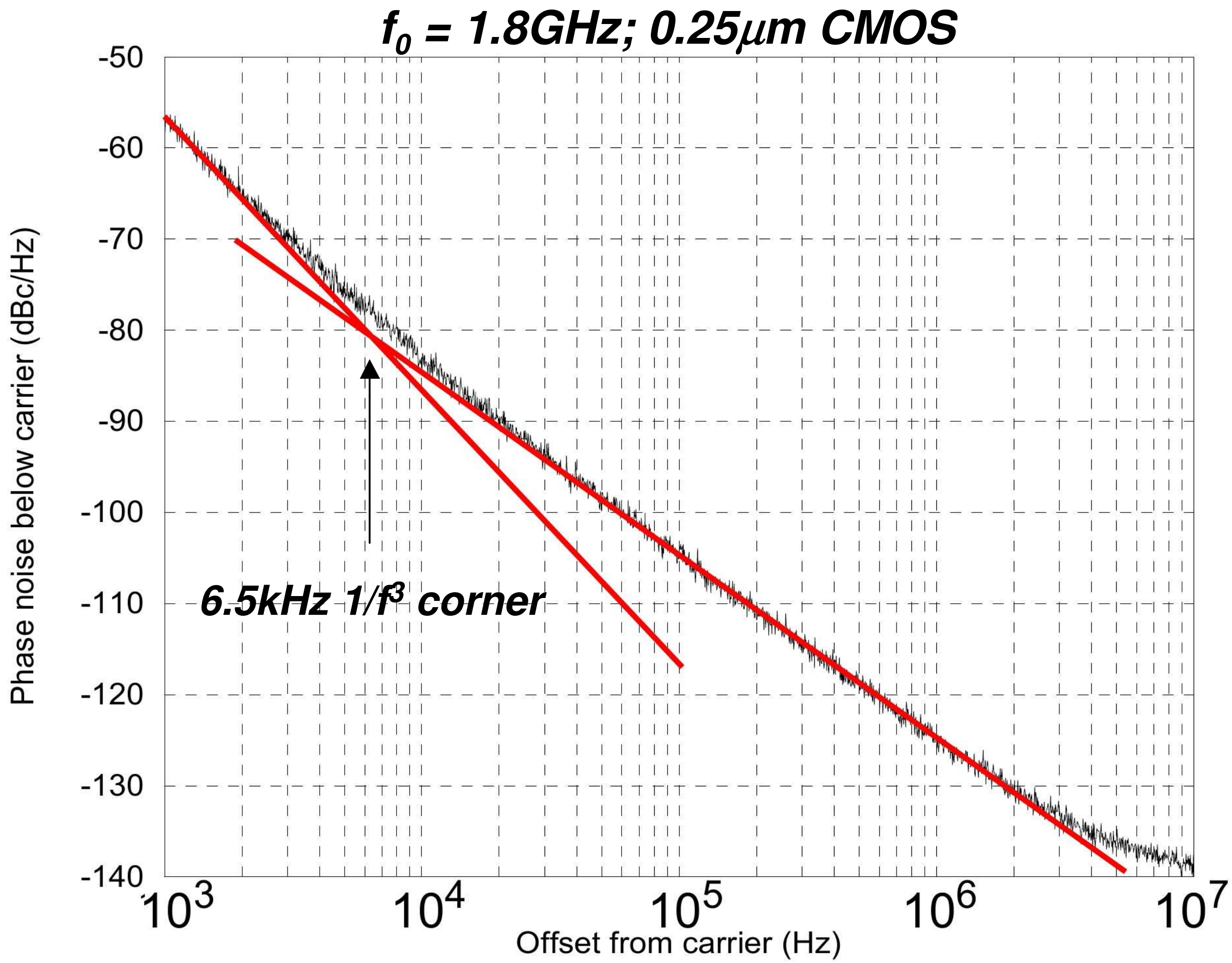


Die photo of complementary oscillator

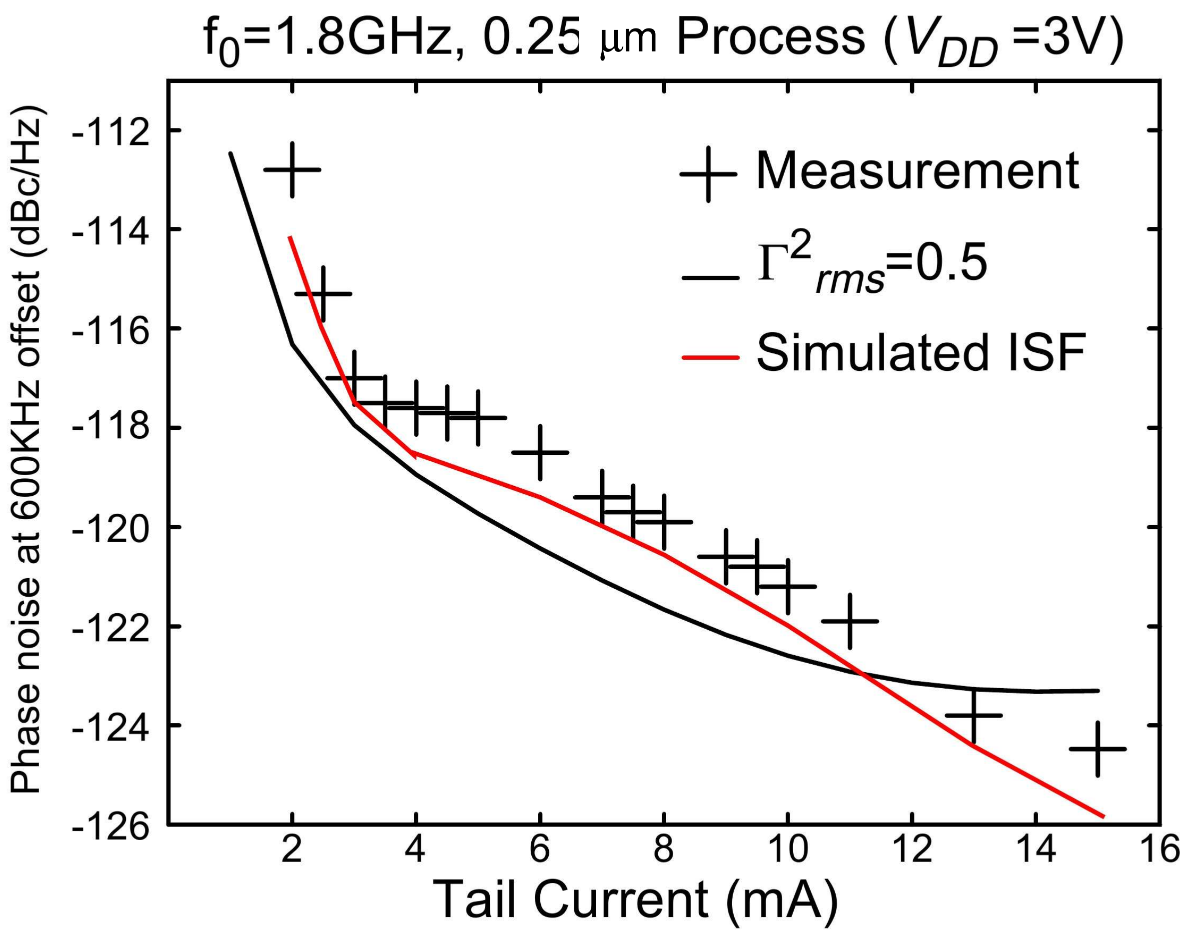


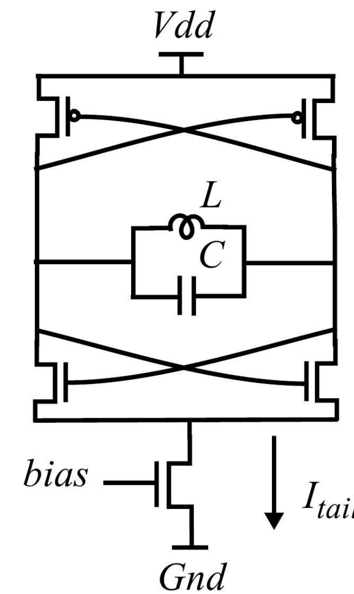
0.25μm CMOS
700μm x 600μm
Pad-limited

Measured phase noise

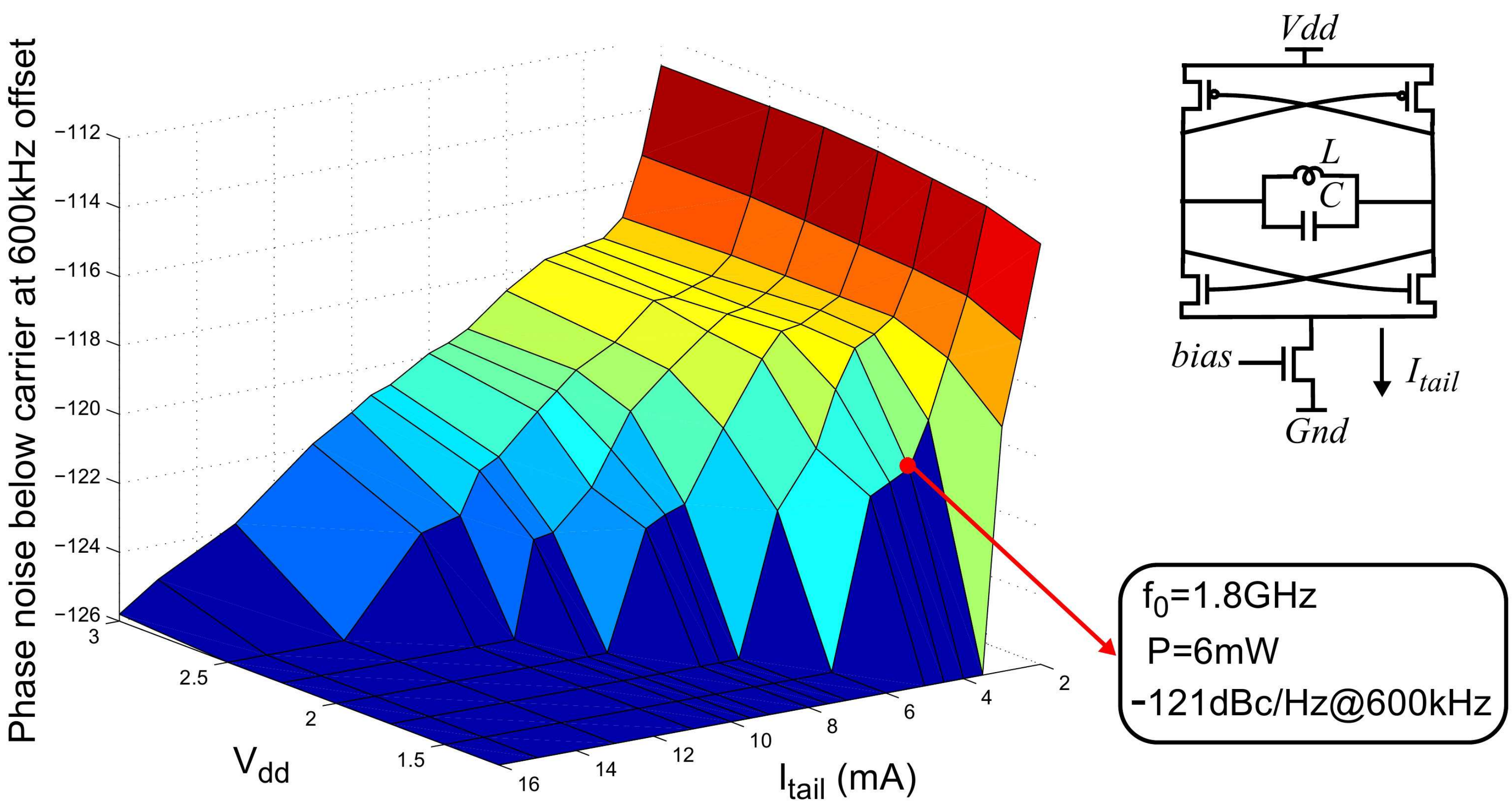


Complementary cross-coupled oscillator

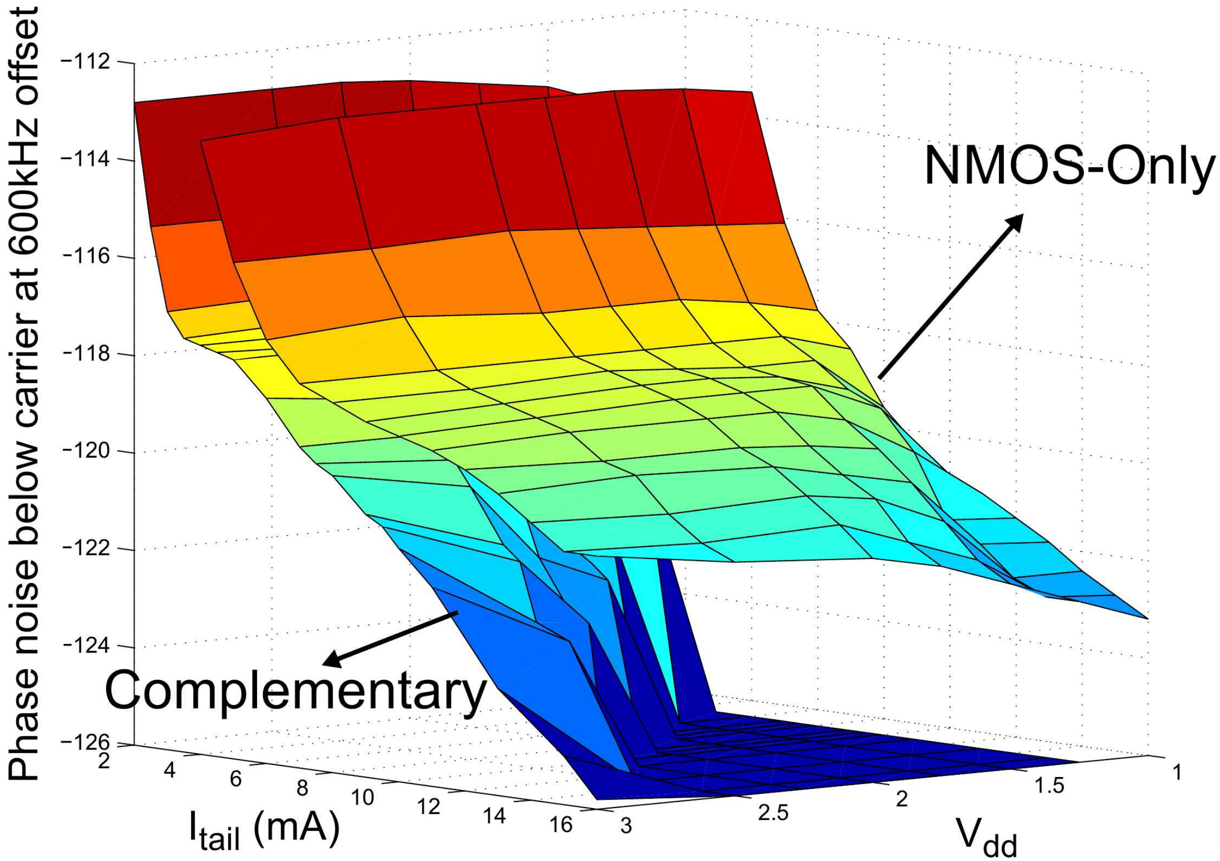


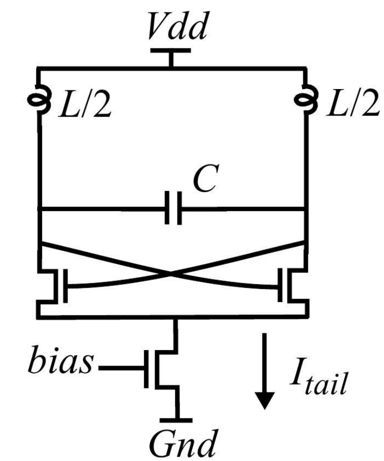


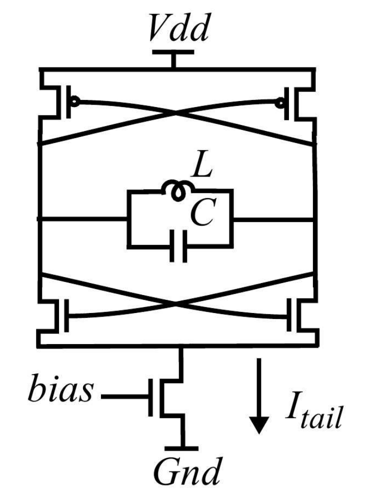
Complementary cross-coupled VCO



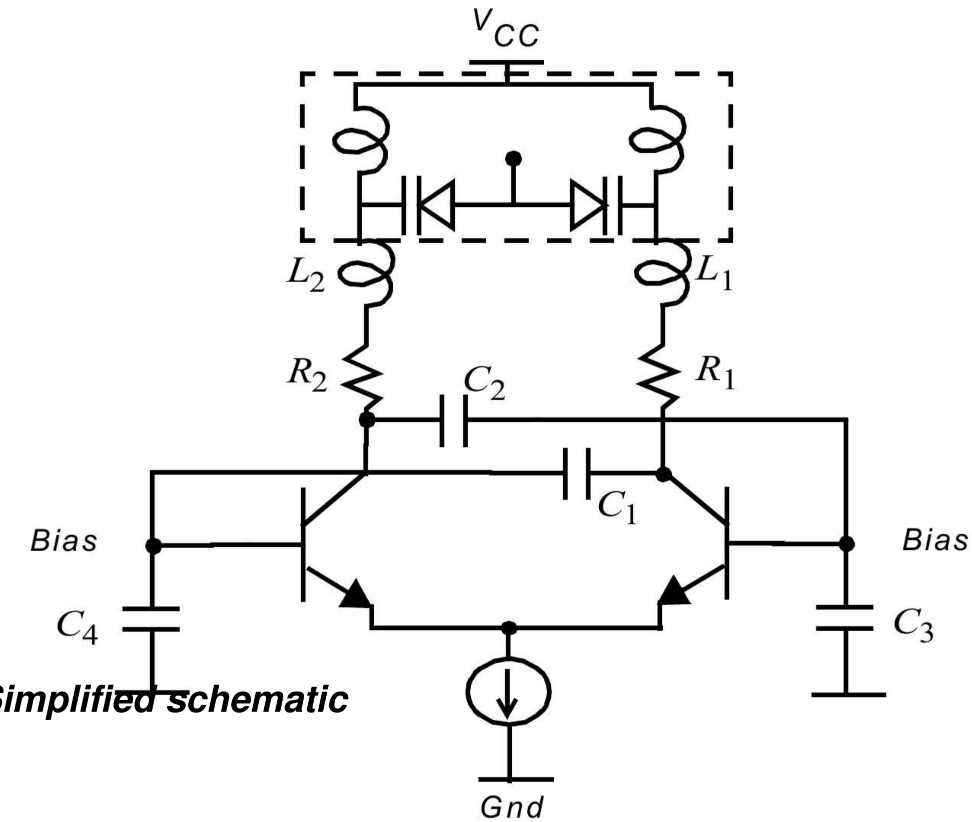
Complementary vs. all-NMOS VCO







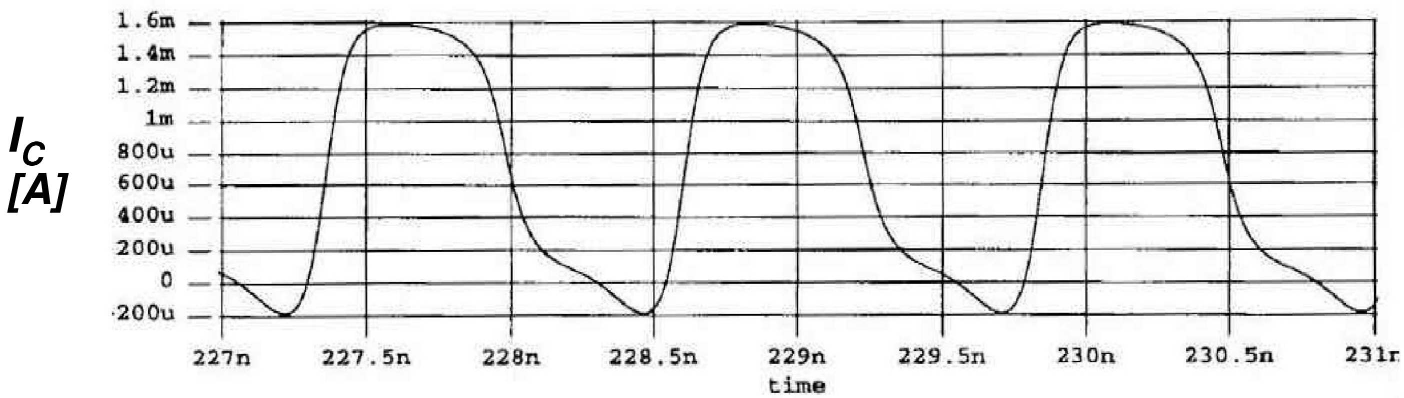
A non-Stanford, non-CMOS example

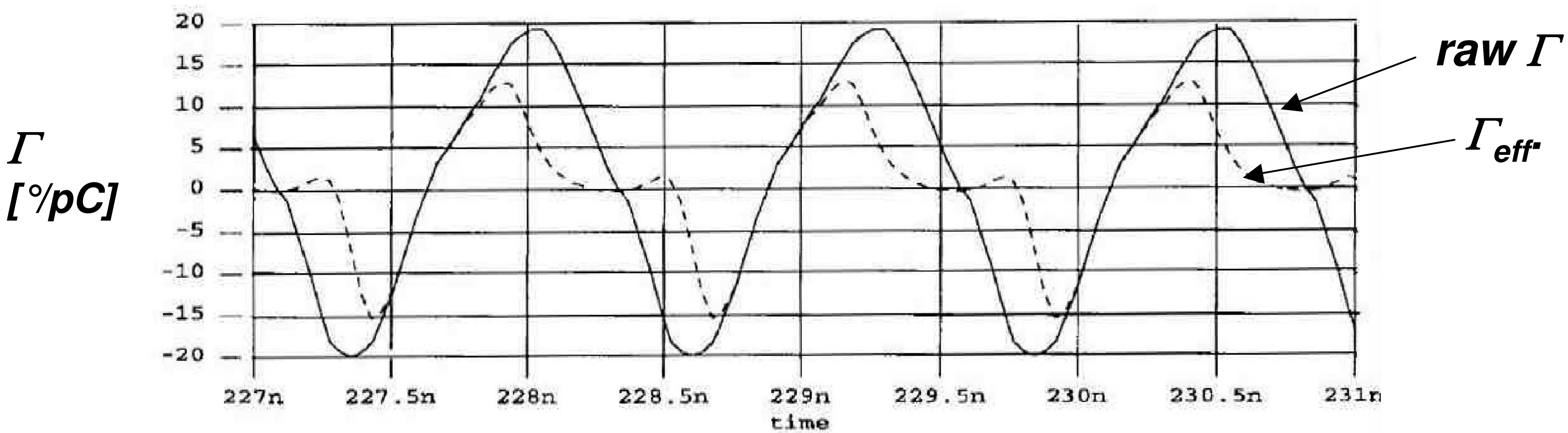


Simplified schematic

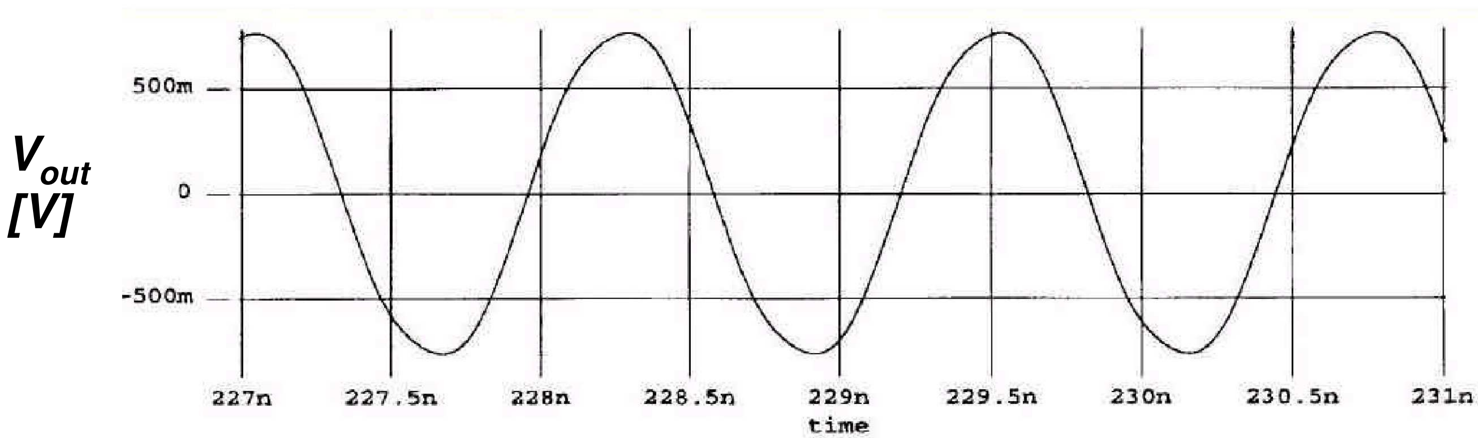
Ref: M.A. Margarit, Joo Leong Tham, R.G. Meyer, M.J. Deen, "A low-noise, low-power VCO with automatic amplitude control for wireless applications," IEEE JSSC, June 1999, pp.761-771.

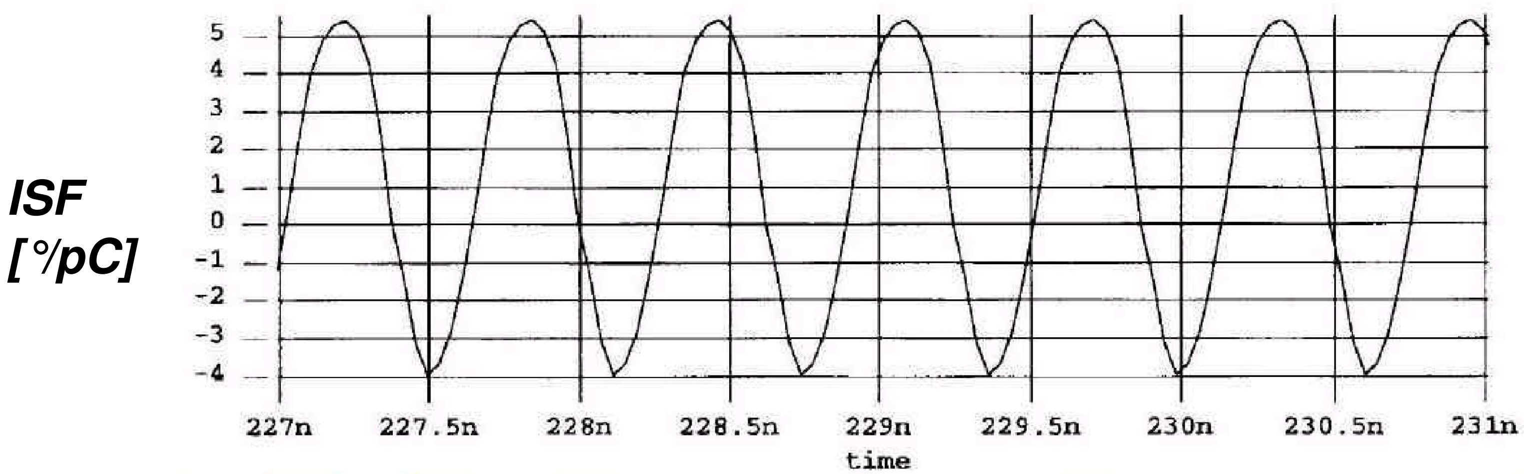
I_C and ISFs for core transistors



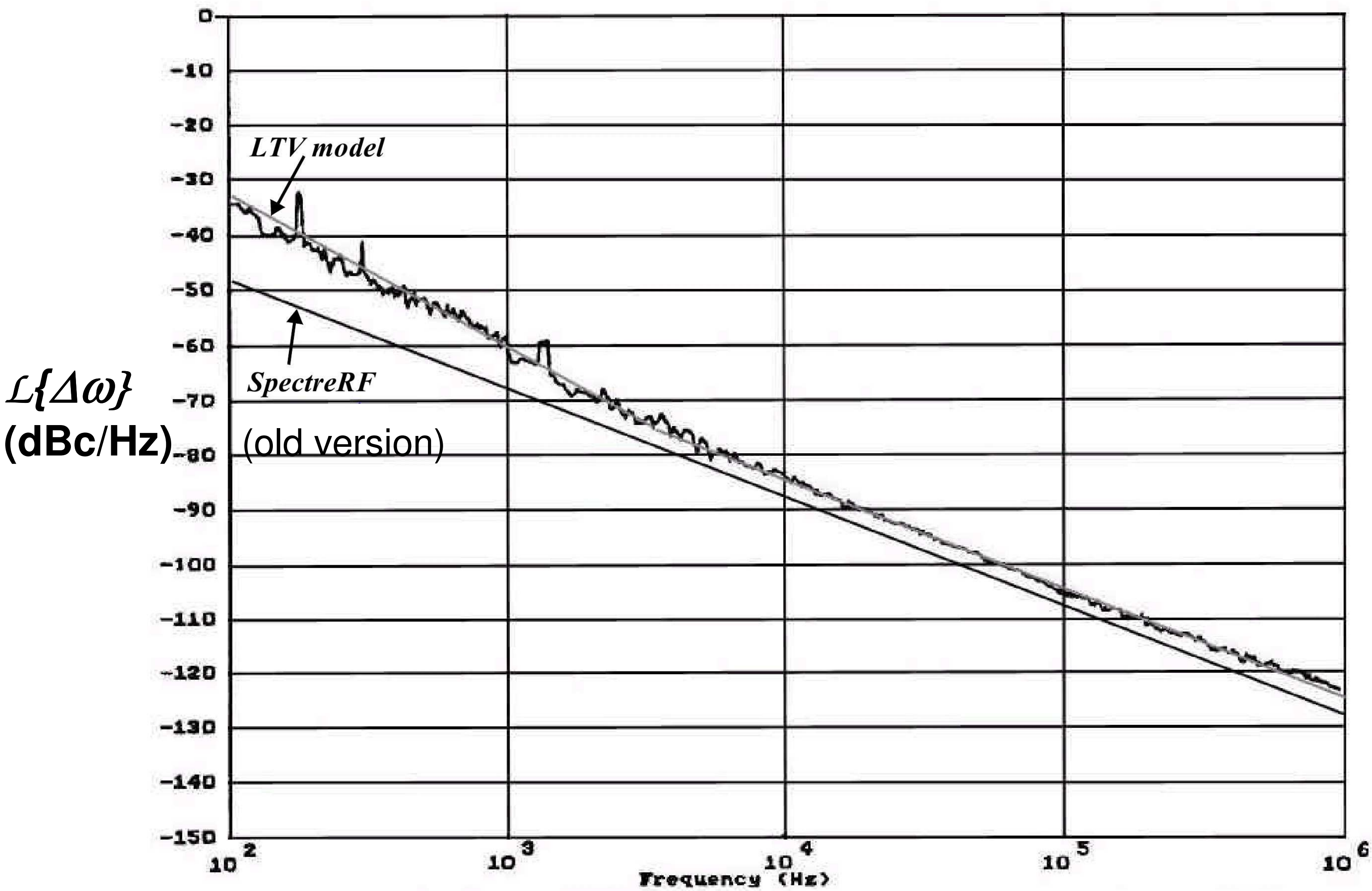


V_{out} and ISF for tail source

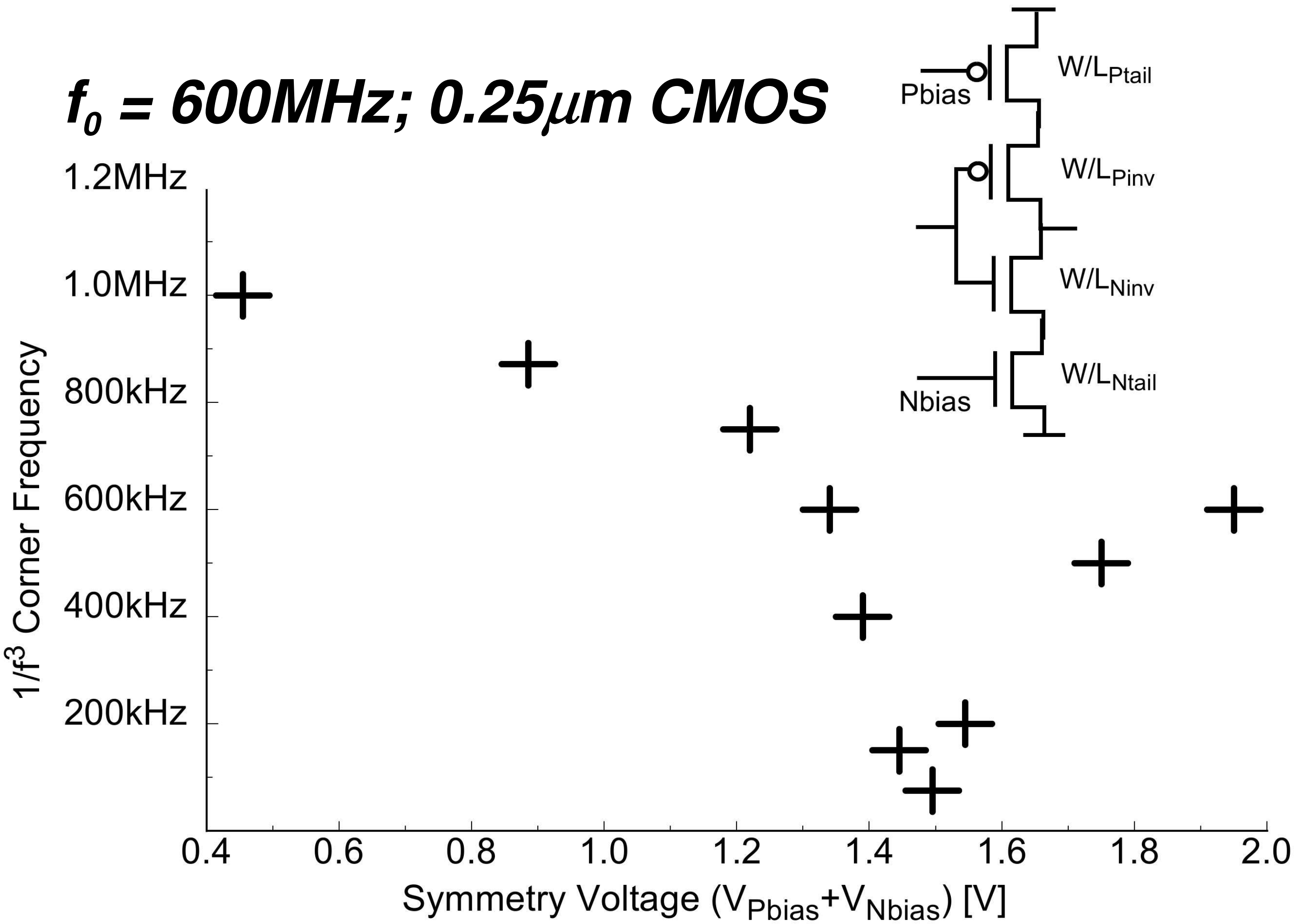




Theory vs. measurement



9-stage current-starved single-ended VCO

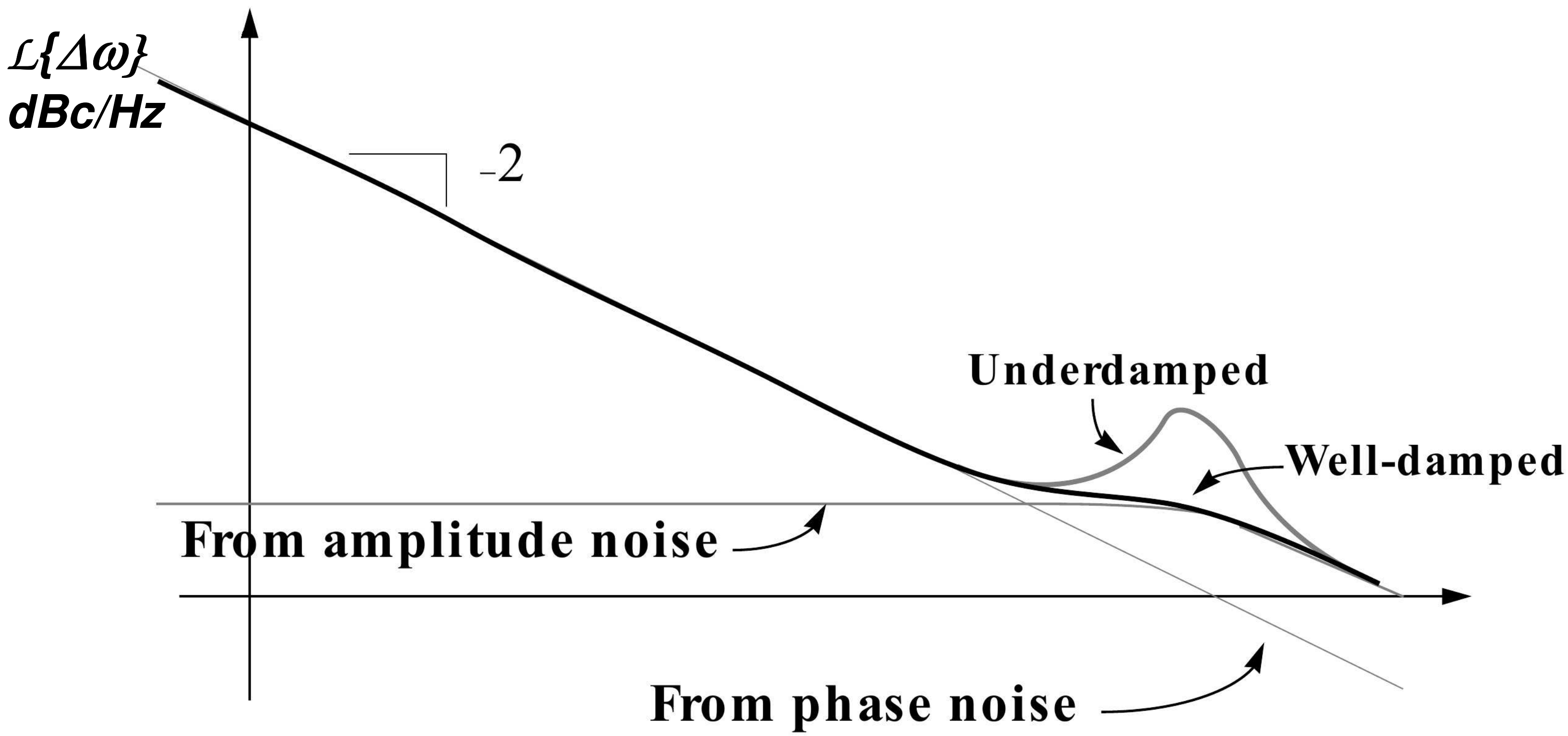


Amplitude noise

- Phase noise generally dominates close-in spectrum. Amplitude noise generally dominates far-out spectrum.
- Amplitude noise can be accommodated with the same impulse-response method.
  - Amplitude control dynamics are frequently approximately single-pole.
    - For an isolated LC tank, the bandwidth is simply ω_0/Q.
    - Contribution to output noise spectrum is flat up to an offset equal to the amplitude-control bandwidth, then rolls off.
    - Superposition with phase noise contribution leads to a characteristic pedestal in the oscillator spectrum.
  - If dynamics are second-order, there can be peaking in the oscillator spectrum.
- Sideband asymmetry can also result from superposition.

Amplitude response

- Possible spectra resulting from different amplitude control dynamics are as follows:



Summary and conclusions

- LTI theories say:
  - Maximize signal power and Q, operate at incipient voltage-limiting, with minimum L/R.
  - One cannot do anything about $1/f$ noise upconversion.
    - $1/f^3$ corner is purely a function of technology and bias.
- LTV theory says:
  - LTI is generally right in the first bullet above.
  - LTI is wrong about the second bullet: Exploiting topological symmetry to minimize the ISF dc value can suppress $1/f$ noise upconversion.

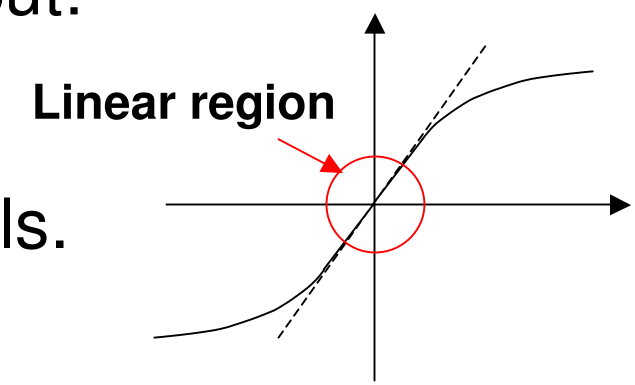
Nonlinearity and Distortion

Linear circuits

- In a linear system the output is always a linear function of the input:

$$v_o = c_1 v_i$$

- In real circuits, this relationship is typically valid only for small signals.



- The output of a nonlinear and memoryless circuit can be approximated with a polynomial:

$$v_o = f(v_{in}) \approx c_0 + c_1 v_i + c_2 v_i^2 + c_3 v_i^3 + \ldots$$

First: *Harmonic* distortion

- Assume the input is a single sinusoid:

$$v_i = V_i \cos(\omega t)$$

- A linear amplifier would only produce:

$$v_o = c_1 V_i \cos(\omega t + \phi)$$ ignore phase today

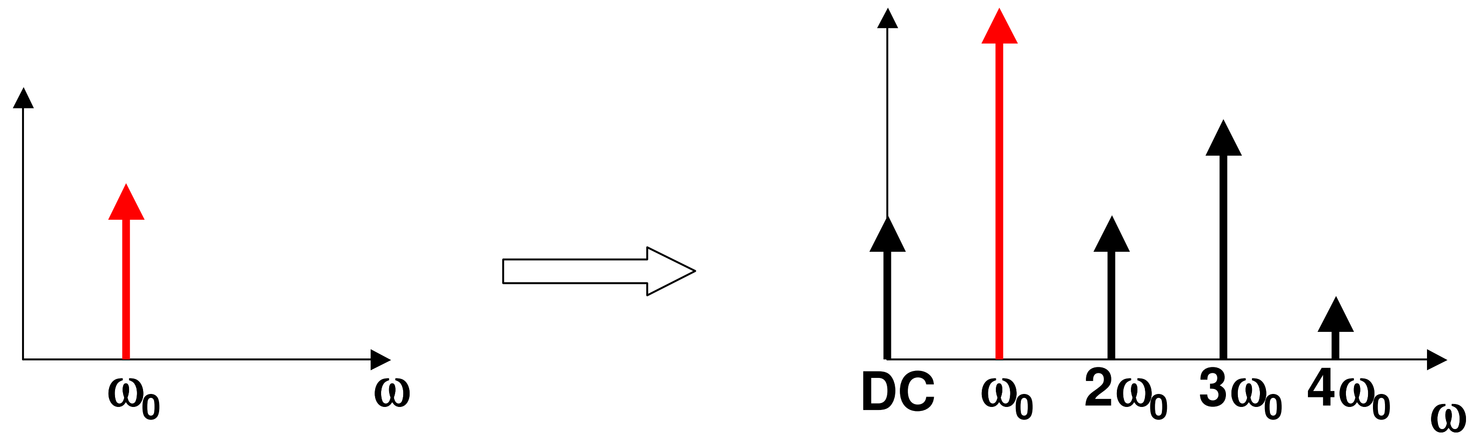
- But the output of a memoryless nonlinear system is:

$$v_o = c_0 + c_1 V_i \cos(\omega t) + c_2 V_i^2 \cos^2(\omega t) + c_3 V_i^3 \cos^3(\omega t) + \ldots$$

$$v_o = c_0 + c_1 V_i \cos(\omega t) + \frac{c_2 V_i^2}{2}[1 + \cos(2\omega t)] + \frac{c_3 V_i^3}{4}[3\cos(\omega t) + \cos(3\omega t)] + \ldots$$

Harmonic distortion (cont)

$$v_o = c_0 + c_1 V_i \cos(\omega t) + \frac{c_2 V_i^2}{2}[1 + \cos(2\omega t)] + \frac{c_3 V_i^3}{4}[3\cos(\omega t) + \cos(3\omega t)] + \ldots$$



- As a result of nonlinearity not only do we get signals at the fundamental frequency (ω), we also get signals at multiples of the input frequency (2ω, 3ω, …).
- Even-order nonlinearities produce even-order harmonics (2ω, 4ω,…), which include DC.
- Odd-order nonlinearities produce odd-order harmonics (3ω, 5ω,…), which include the fundamental frequency (ω).

Fractional harmonic distortion

- Fractional second-harmonic distortion is a normalized measure:

$$HD2 = \frac{\text{amplitude of the second harmonic}}{\text{amplitude of the fundamental}}$$

- If we assume the squared term dominates the second harmonic

$$\boxed{HD2 = \frac{\frac{1}{2}c_2 V_i^2}{c_1 V_i} = \frac{1}{2}\frac{c_2}{c_1}V_i}$$

- HD2 is proportional to input amplitude.
- Similarly HD3 is defined as

$$\boxed{HD3 = \frac{\text{amplitude of the third harmonic}}{\text{amplitude of the fundamental}} = \frac{\frac{1}{4}c_3 V_i^3}{c_1 V_i} = \frac{c_3}{4c_1}V_i^2}$$

- HD3 is proportional to the input power.

Total harmonic distortion

- Since the harmonics are orthogonal the total output power is the sum of the harmonic powers.
- Total harmonic distortion (THD) is defined as:

$$THD = \sqrt{\frac{\text{power in distortion}}{\text{power in fundamental}}}$$

- Therefore:

$$\boxed{THD = \sqrt{HD2^2 + HD3^2 + \ldots}}$$

Intermodulation distortion

- So far we have characterized a nonlinear system using a single input tone. Now consider two tones:

$$v_i = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t$$

$$v_o = c_0 + c_1 v_i + c_2 v_i^2 + c_3 v_i^3 + \ldots$$

- The squared term gives:

$$c_2[V_1^2 \cos^2 \omega_1 t + V_2^2 \cos^2 \omega_2 t + 2V_1 V_2 \cos \omega_1 t \cos \omega_2 t]$$

$$= c_2 \frac{V_1^2}{2}(1 + \cos 2\omega_1 t) + c_2 \frac{V_2^2}{2}(1 + \cos 2\omega_2 t) +$$

$$c_2 V_1 V_2 [\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t]$$

Second-order intermodulation

- The last term, $\cos(\omega_1 \pm \omega_2)t$, is the second-order intermodulation (IM) product.
- The normalized second-order IM distortion is defined with two input tones of equal amplitude:

$$V_1 = V_2 = V_i$$

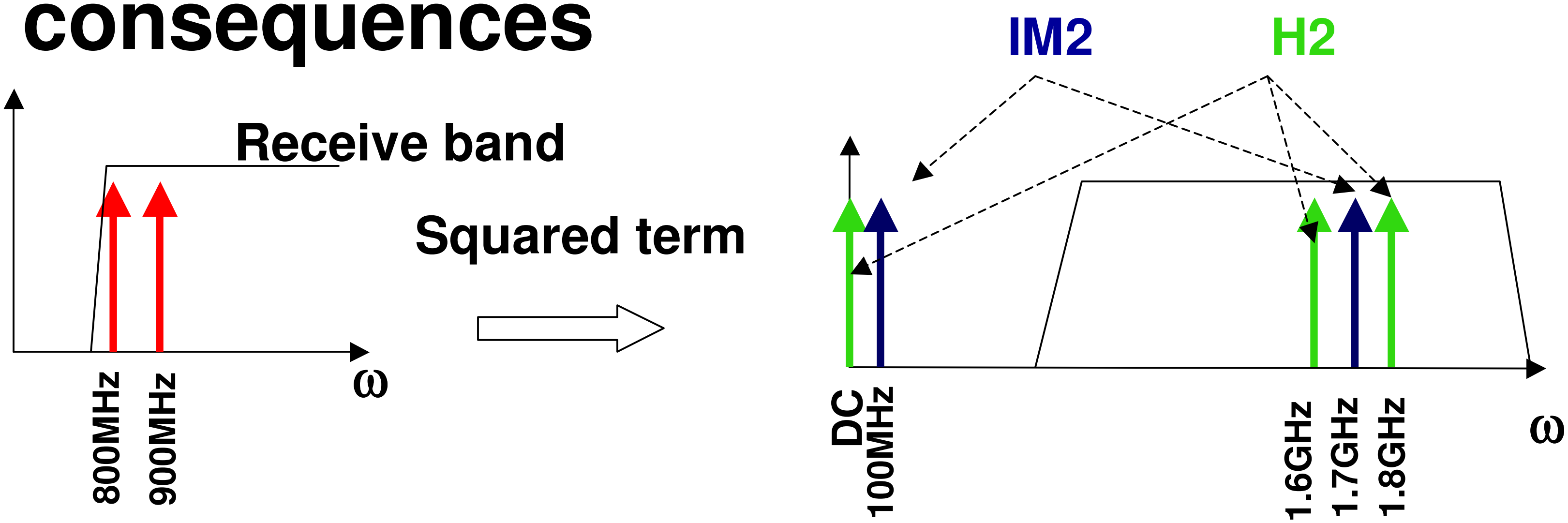
$$IM2 = \frac{\text{amplitude of the intermod}}{\text{amplitude of the fundamental}} = \frac{c_2}{c_1} V_i$$

- Note that

$$IM2 = 2HD2$$

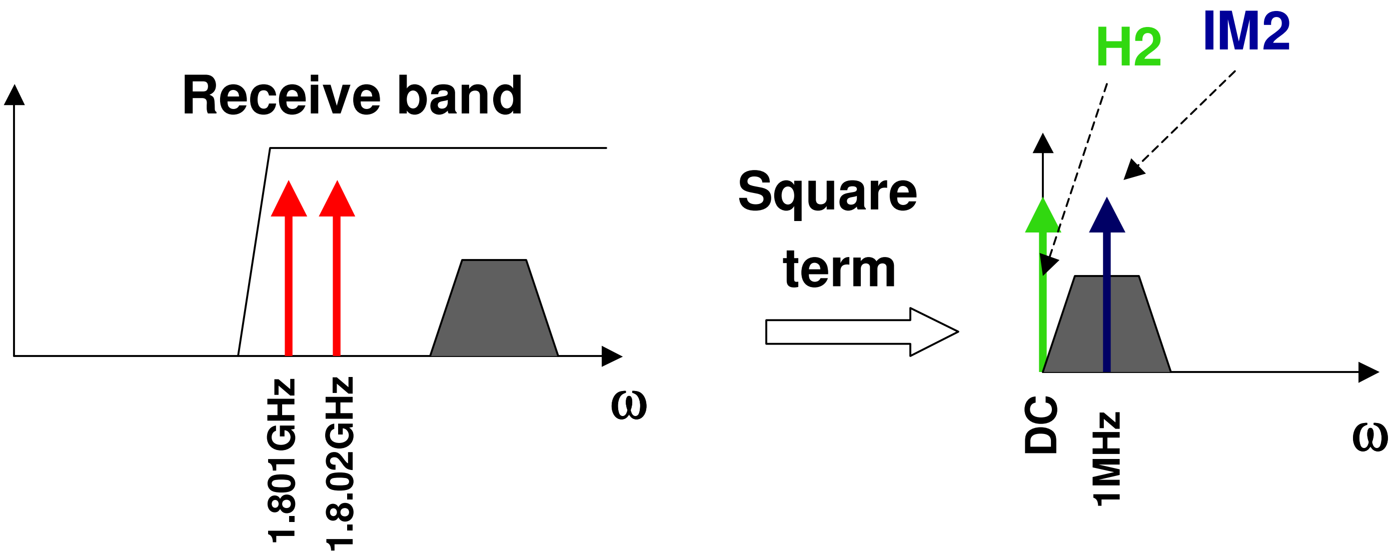
$$IM2\big|_{dB} = HD2\big|_{dB} + 6dB$$

IM2 consequences



- IM2 produces both low frequency (ω_1-ω_2) and high frequency (ω_1+ω_2) components.
- The term at (ω_1+ω_2) can be a problem in wideband receivers (i.e., those with bandwidth larger than one octave).
- Example: Suppose the receiver covers 800MHz-2.4GHz and two interfering signals appear at 800MHz and 900MHz.
- The 2nd-order IM terms are at 100MHz and 1.7GHz; the second harmonics are at 1.6GHz and 1.8GHz. Thus IM2 and HD2 generate three components within the receive band.

IM2 issues for low-IF and zero-IF receivers



- The low frequency of the second-order IM term (ω_1- ω_2) can be a problem in low-IF or zero-IF receivers
- Example: Assume a receiver covers 1.8GHz-1.9GHz and has an IF=1MHz. Now assume two interfering signals at 1.801MHz and 1.802MHz.
- The second-order IM distortion product is at the IF frequency and therefore constitutes potentially serious interference.

Third-order intermodulation

- Now if we consider the cubic term we get:

$$c_3 v_i^3 = c_3[V_1 \cos \omega_1 t + V_2 \cos \omega_2 t]^3$$

- Expansion yields, in part, terms which are the same as the cubic distortion with a single input:

$$\frac{c_3 V_{1,2}^3}{4}[3\cos \omega_{1,2} t + \cos 3\omega_{1,2} t]$$

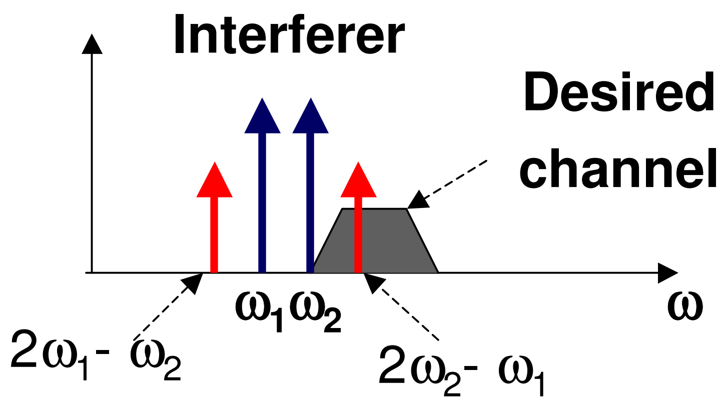
- But we *also* get cross terms:

$$3c_3[V_1 V_2^2 \cos \omega_1 t \underbrace{\cos^2 \omega_2 t}_{\frac{1}{2}[1+\cos 2\omega_2 t]} + V_1^2 V_2 \underbrace{\cos^2 \omega_1 t}_{\frac{1}{2}[1+\cos 2\omega_1 t]} \cos \omega_2 t]$$

$$= 3c_3 (V_1 V_2^2 [\frac{1}{2}\cos \omega_1 t + \frac{1}{4}\cos(2\omega_2 \pm \omega_1)t] + V_1^2 V_2 [\frac{1}{2}\cos \omega_2 t + \frac{1}{4}\cos(2\omega_1 \pm \omega_2)t])$$

Third-order intermodulation (cont.)

$$3c_3(V_1V_2^2[\frac{1}{2}\cos\omega_1 t+\frac{1}{4}\cos(2\omega_2\pm\omega_1)t]+V_1^2V_2[\frac{1}{2}\cos\omega_2 t+\frac{1}{4}\cos(2\omega_1\pm\omega_2)t])$$



- If $\omega_1 \approx \omega_2$ then $2\omega_1 - \omega_2 \approx 2\omega_2 - \omega_1 \approx \omega_1 \approx \omega_2$.
- Unlike a 2nd-order nonlinearity, which creates in-band IM only if the system is wideband, a cubic nonlinearity can create in-band IM even in narrowband systems.

IM3 definition

- The $\cos(2\omega_1 \pm \omega_2)t$ and $\cos(2\omega_2 \pm \omega_1)t$ terms are third-order intermodulation distortion terms.
- The normalized third-order IM is defined with two input tones of equal amplitude:

$$V_1 = V_2 = V_i$$

$$IM3 = \frac{\text{amplitude of the intermod}}{\text{amplitude of the fundamental}} = \frac{3}{4}\frac{c_3}{c_1}V_i^2$$

- The following relationships hold:

$$IM3 = 3HD3; \qquad IM3\big|_{dB} = HD3\big|_{dB} + 9.5dB$$

IP2

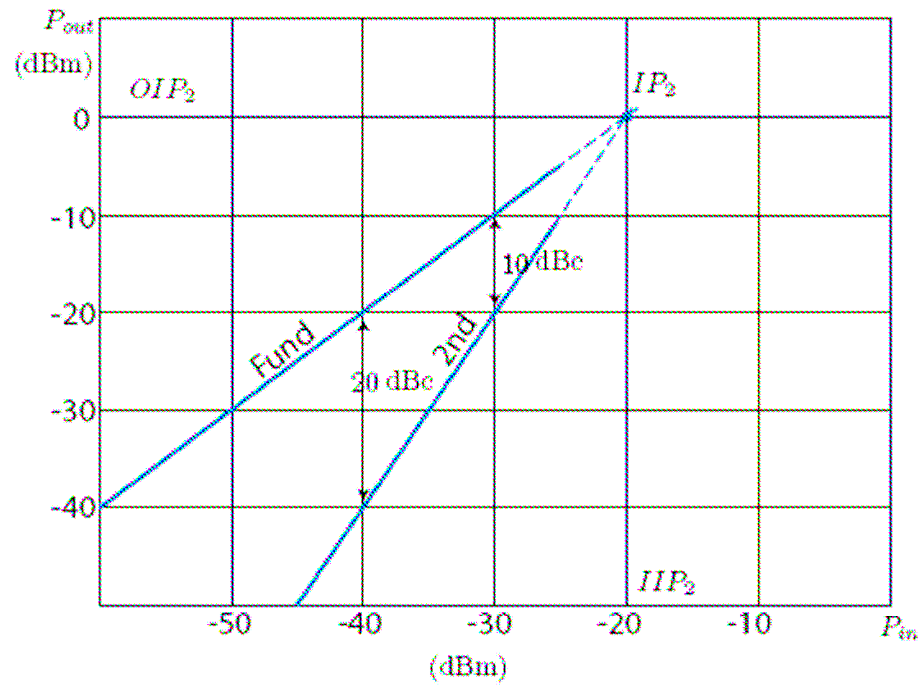
- The fundamental term increases linearly with the input signal:

$$V_o = c_1 V_i$$

  - On a log scale the input-output curve has unit slope.

- The second-order IM term is proportional to the square of the input signal

$$V_o = c_2 V_i^2$$



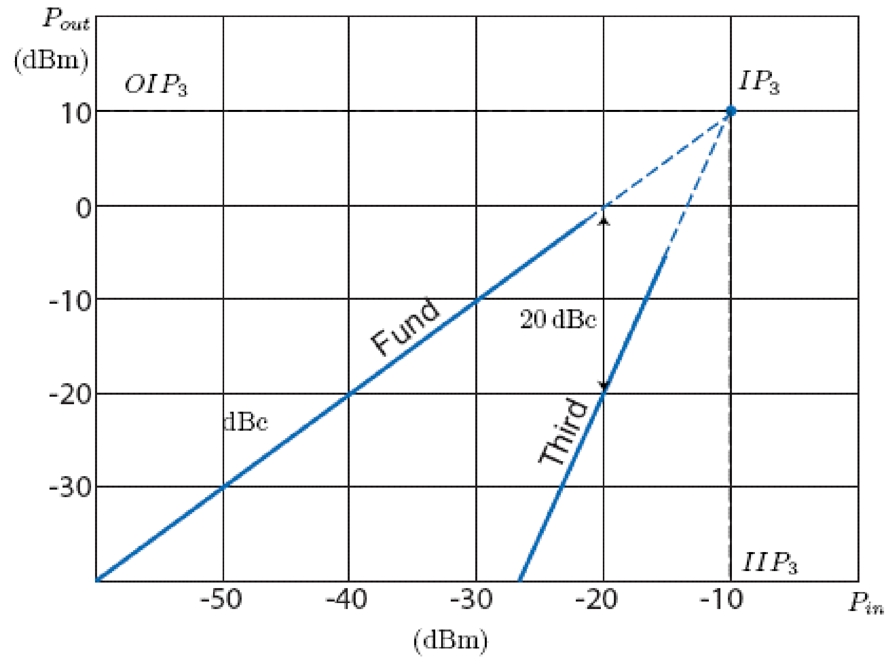
  - On a log scale the input-intermod curve thus has a slope of 2.

- IP2 is the **extrapolated** intercept point between the fundamental term and the second-order IM term.
  - *At* IP2 we have IM2=0dBc
  - When IP2 is known we can back-calculate IM2 at any other power.
    - For each dB input power drop, IM2 drops a dB:

$$IM2(dB) = P_{in}(dB) - IIP2(dB)$$

IP3

- The fundamental term (still) increases linearly with input signal.
- The third-order IM term is proportional to the cube of the input signal: $$V_o = \frac{3}{4} c_3 V_i^3$$
  - On a log scale the input-intermod curve thus has a slope of 3.



- IP3 is the **extrapolated** intercept point between the fundamental term and the third-order IM term.
  - At IP3 we have IM3=0dBc
  - When IP3 is known we can back-calculate IM3 at any other power.
    - For each dB input power drop, IM3 drops by 2dB:

$$IM3(dB) = 2[P_{in}(dB) - IIP3(dB)]$$

IP3 (cont.)

- We can calculate the IP3 from the cubic non-linearity coefficient:

$$V_o\Big|_{IP3} = \left|\frac{3}{4}c_3 V_i^3\right| = \left|c_1 V_i\right| \Rightarrow V_{i,ip3} = \sqrt{\frac{4}{3}\left|\frac{c_1}{c_3}\right|}$$

- Notice that this calculation is correct even if other higher-order nonlinearities exist.
  - Remember IP3 is by definition an *extrapolation*. If we look at small enough signal levels, the cubic term would dominate the odd-order nonlinearity.

Example 1

- Suppose an amplifier has a small-signal power gain of 12dB, OIP3=13dBm and OIP2=30dBm. What are IM2 and IM3 at Pin=-10dBm and Pin=-100dBm?

$$IIP2 = OIP2 - G = 30 - 12 = 18\text{dBm}$$

$$IIP3 = OIP3 - G = 13 - 12 = 1\text{dBm}$$

$$IM2\big|_{-10\text{dBm}} = -10 - IIP2 = -28\text{dBc}$$

$$IM3\big|_{-10\text{dBm}} = 2*(-10 - IIP3) = -22\text{dBc}$$

$$IM2\big|_{-100\text{dBm}} = -100 - IIP2 = -118\text{dBc}$$

$$IM3\big|_{-100\text{dBm}} = 2*(-100 - IIP3) = -202\text{dBc}$$

Example 2

- In the previous example what is THD at Pin=-10dBm? Assume the amplifier has only square and cubic nonlinearities.

$$IM2 = HD2 + 6 \Rightarrow HD2\Big|_{-10\text{dBm}} = -28 - 6 = -34\text{dB} = 10^{-34/20} = 0.02$$

$$IM3 = HD3 + 9.5 \Rightarrow HD3\Big|_{-10\text{dBm}} = -22 - 9.5 = -31.5\text{dB} = 10^{-31.5/20} = 0.027$$

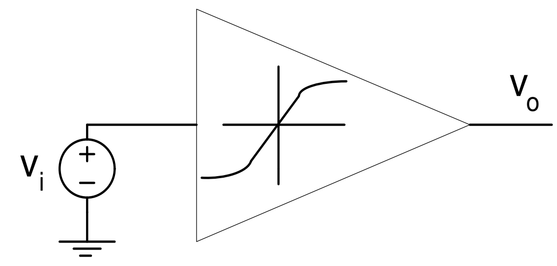
$$THD = \sqrt{HD2^2 + HD3^2} = \sqrt{0.02^2 + 0.027^2} = 0.034 = -29.4\text{dBc}$$

IP3 measurement

- We can measure IP3 with a single two-tone test!
- Let

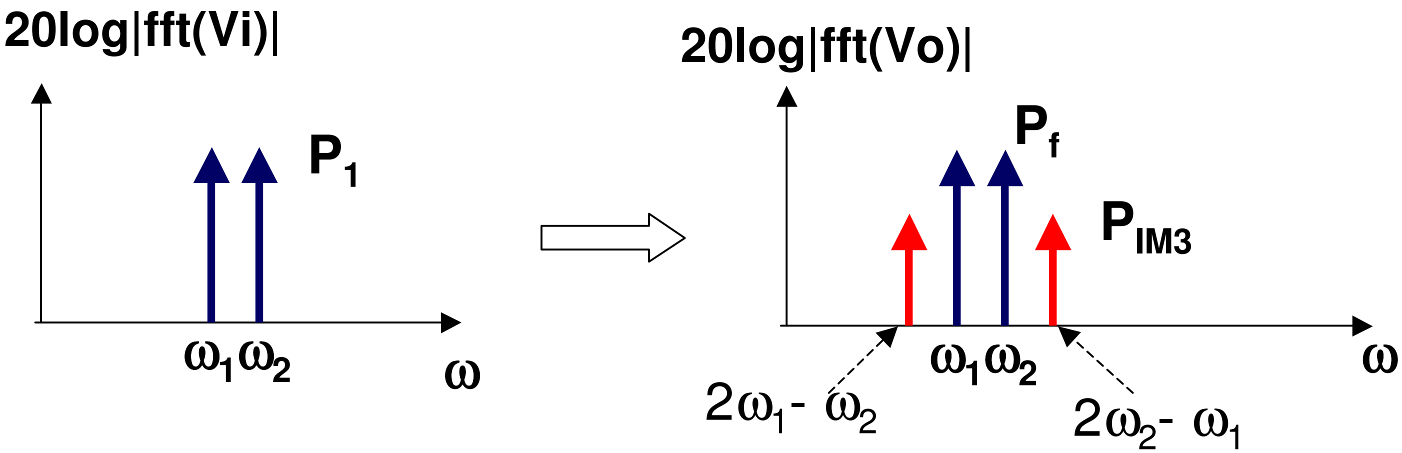
$$v_i = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t$$

$$V = V_1 = V_2$$

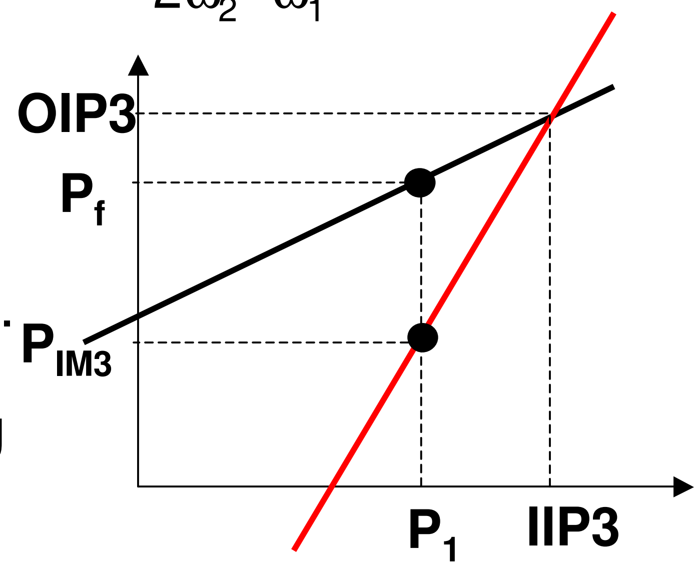


- Choose $\omega_1 \sim \omega_2$ in the middle of the operating frequency range.
- Now can run a transient simulation and take the FFT of the output to get:

IP3 measurement (cont.)



- Now we can use P_1, P_f and P_{IM3} to estimate IP3.
- Remember the fundamental output is linear in input power, so that the log Pout-Pin curve is a line with unit slope.
- The IM3 term is proportional to the cube of the input power so that the log PIM3-Pin curve is a line with slope of three.



IP3 measurement (cont.)

- Now we can write the line equations for the fundamental and IM3 terms:

$$P_o = P_f + (P_i - P_1)$$

$$P_{o,IM3} = P_{IM3} + 3(P_i - P_1)$$

- IP3 is defined at $P_o = P_{o,IM3}$, therefore

$$OIP3 = P_f + (IIP3 - P_1) = P_{IM3} + 3(IIP3 - P_1)$$

$$IIP3 = \frac{P_f - P_{IM3} + 2P_1}{2}$$

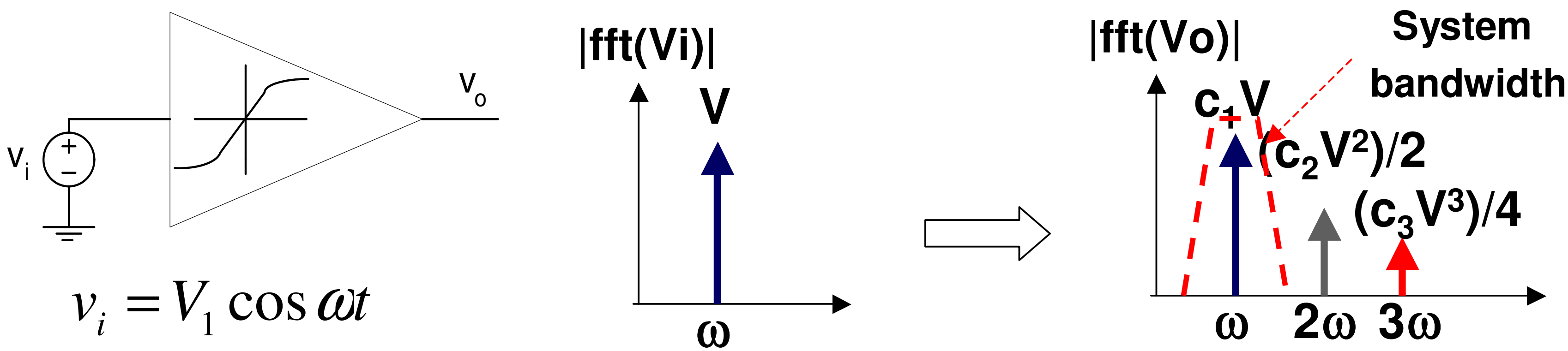
$$OIP3 = \frac{3P_f - P_{IM3}}{2}$$

IP3 estimation from harmonic measurement

- Remember from previous slides that:

$$V_{i,ip3} = \sqrt{\frac{4}{3}\left|\frac{c_1}{c_3}\right|}$$

- Then we should be able to use just a single tone and measure c1 and c3 by measuring the fundamental and 3rd harmonic:



$$v_i = V_1 \cos \omega t$$

- This technique works only for systems where the 3rd harmonic falls in band.

Example 3

- You have designed an amplifier with 20dB power gain at 2GHz. The input impedance is 50Ω and the output impedance is 200Ω. Your test with 10mV tones at 1.99GHz and 2.01GHz results in an IM3 of 100μV at the output. What are the IIP3 and OIP3 of this amplifier?
- We first need to calculate the amplitude of the fundamental signal at the output:

$$G_p = \frac{P_o}{P_i} = \frac{V_o^2 / 2R_o}{V_i^2 / 2R_i} = \frac{V_o^2}{V_i^2}\frac{R_i}{R_o} = G_v^2 \frac{R_i}{R_o} \Rightarrow G_v = \sqrt{G_p \frac{R_o}{R_i}} = \sqrt{100\frac{200}{50}} = 20$$

$$V_{o,f} = G_v V_i = 20 * 10mV = 200mV$$

- Then IIP3 is:

$$IIP3 = \frac{P_f - P_{IM3} + 2P_i}{2} = \frac{20\log(\frac{200mV}{100\mu V}(10mV)^2)}{2} = -6.9dBv = 0.45V$$

$$OIP3 = IIP3 * G_v = 0.45 * 20 = 9V$$

Example 3 (cont)

- If the amplifier's 3dB bandwidth is 200MHz, then what would be HD3 for a 10mV single tone input?
- Let's first ignore the amplifier bandwidth:

$$IM3(dB) = 2[P_{in}(dB) - IIP3(dB)] = 2*20\log(10mV/0.45) = -66dBc$$

$$HD3 = IM3 - 9.5dB = -75.5dBc$$

- If we assume an *RLC* tank at the output,

$$BW = \frac{\omega_0}{Q} \Rightarrow Q = \frac{2GHz}{200MHz} = 10$$

Example 3 (cont.)

- Now we can estimate HD3 after filtering

$$HD3 = HD3\Big|_{\text{w/o filter}} + 10\log\left(\frac{1}{1+[Q(\frac{3\omega}{\omega}-\frac{\omega}{3\omega})]^2}\right)$$

$$= -75.5 - 28.5 = -104dBc$$

Gain at fundamental frequency

- Let's revisit the single-tone excitation case:

$$v_o = c_0 + c_1 V_i \cos(\omega t) + c_2 V_i^2 \cos^2(\omega t) + c_3 V_i^3 \cos^3(\omega t) + \ldots$$

$$v_o = c_0 + c_1 V_i \cos(\omega t) + \frac{c_2 V_i^2}{2}[1 + \cos(2\omega t)] + \frac{c_3 V_i^3}{4}[3\cos(\omega t) + \cos(3\omega t)] + \ldots$$

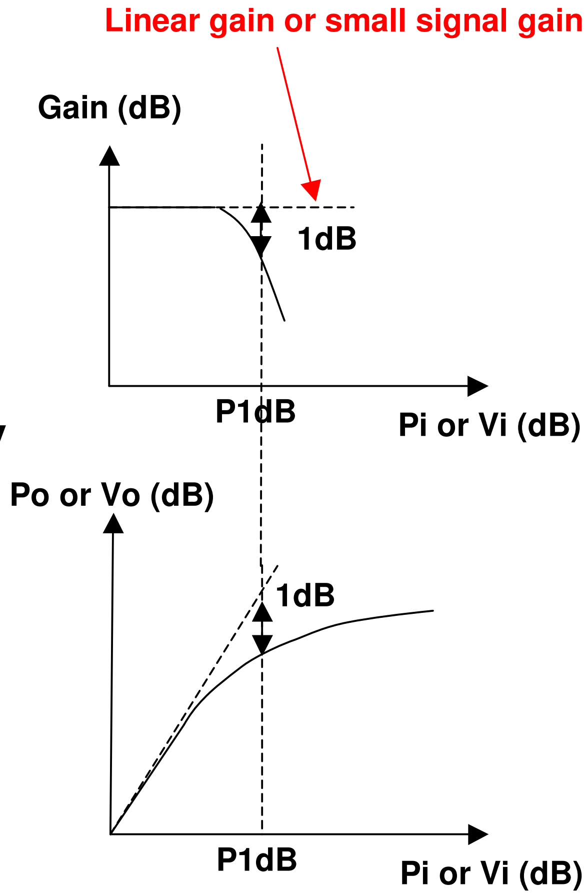
$$v_o = [c_0 + \frac{c_2 V_i^2}{2}] + \underbrace{[c_1 + \frac{c_3 V_i^2}{4}]}_{\text{gain at } \omega} V_i \cos(\omega t) + \frac{c_2 V_i^2}{2}\cos(2\omega t) + \frac{c_3 V_i^3}{4}\cos(3\omega t) + \ldots$$

- It appears that gain at the fundamental frequency is now a function of the signal level!
- Depending upon the polarity of c_3 we can get gain *compression* or gain *expansion* (peaking) at the fundamental frequency.

$$\boxed{\left.\frac{v_o}{v_i}\right|_{\omega} = c_1 + \frac{3V_i^2}{4}c_3}$$

1dB compression point

- The 1dB compression point by definition is the power/signal level at which the large signal gain (i.e., Po/Pi) at the fundamental frequency is reduced by 1dB.
- Remember that only odd-order distortions result in gain change at the fundamental frequency.



1dB compression point (cont.)

- If cubic nonlinearity is the only odd-order nonlinear term then we can calculate the 1dB compression point:

$$20\log(c_1 V_i + \frac{3}{4} c_3 V_i^3) = 20\log(c_1 V_i) - 1dB$$

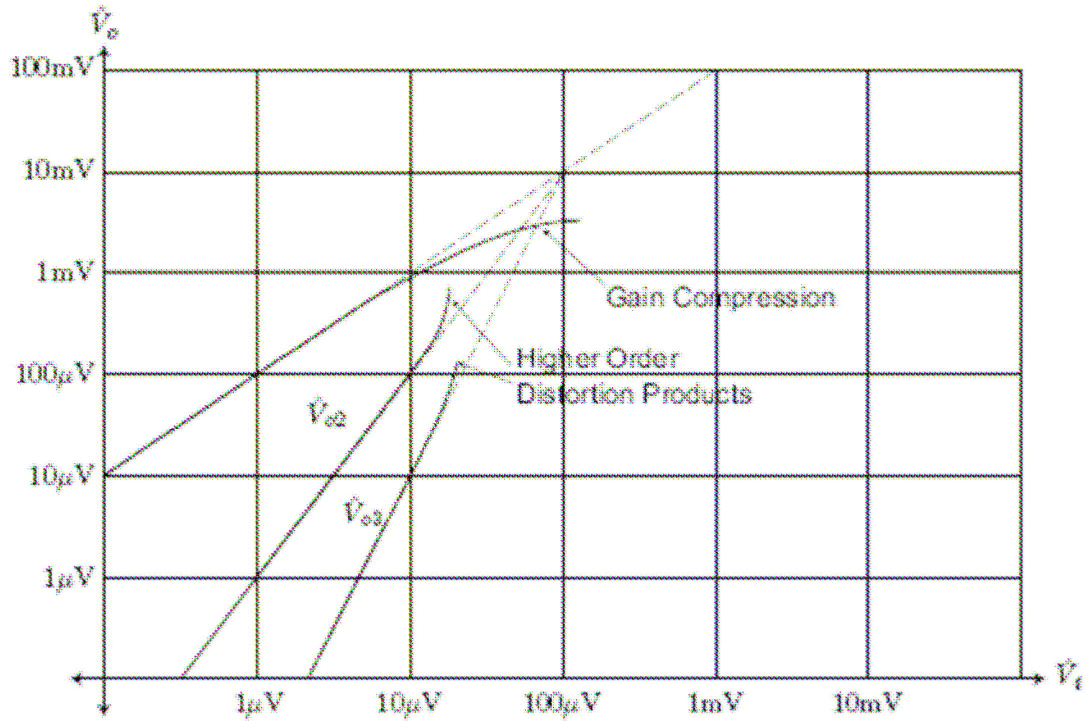
$$1 + \frac{3}{4}\frac{c_3}{c_1} V_i^2 = 0.89 \Rightarrow V_{i,1dB} = \sqrt{0.145\left|\frac{c_1}{c_3}\right|}$$

$$\text{P1dB} = 10\log\left(\left|\frac{c_1}{c_3}\right|\right) + 10\log(0.145) = IP3 + 10\log(0.145 * 3/4)$$

$$\boxed{\text{P1dB} = IP3 - 9.6dB}$$

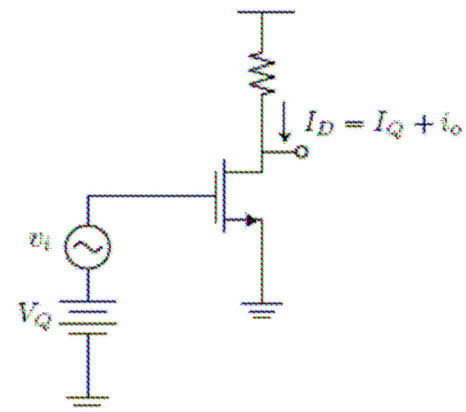
- This relationship is accurate when there is only cubic nonlinearity, as is perhaps explained best graphically!

Distortion



- The higher-order nonlinearities do not affect IP2 and IP3 points, but they control the 1dB compression point.

Distortion in long-channel MOS devices



- If we ignore the output impedance:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L_{eff}}(V_{GS} - V_{th})^2 = \frac{K}{2}(V_{GS} - V_{th})^2$$

$$I_D = \frac{K}{2}(V_Q + v_i - V_{th})^2 = \underbrace{\frac{K}{2}(V_Q - V_{th})^2}_{c_0} + \underbrace{K(V_Q - V_{th})}_{c_1}v_i + \underbrace{\frac{K}{2}}_{c_2}v_i^2$$

- It seems that in long-channel devices IM3=0
  - This cannot be true (because Murphy always wins); what are we missing?

IM3 in "square-law" devices

- One factor we ignored is mobility degradation due to vertical field:

$$\mu_{eff} = \frac{\mu}{1+\theta . V_{OD}}$$

$$\theta \approx \frac{2\times 10^{-9}}{t_{ox}} \text{ V}^{-1} : \text{fitting parameter}$$

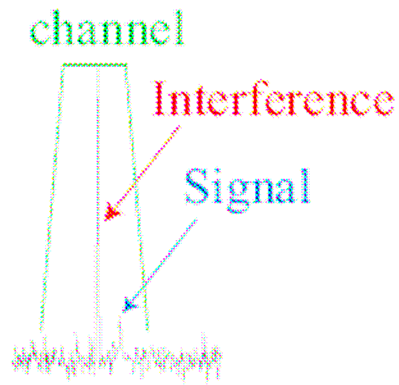
$$\theta \approx 0.7 \text{ V}^{-1} \text{for 90nm CMOS process}$$

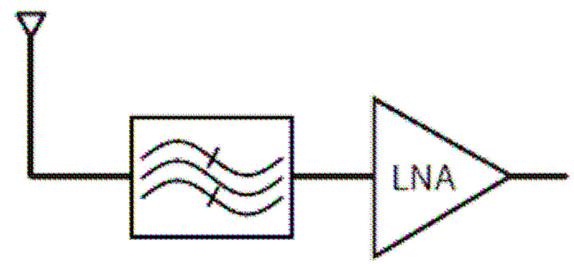
$$V_{OD} = V_{GS} - V_{th}$$

$$I_D = \frac{1}{2}\mu_{eff} C_{ox} \frac{W}{L_{eff}} V_{OD}{}^2 = \frac{K}{2}\frac{V_{OD}{}^2}{1+\theta . V_{OD}}$$

$$\xrightarrow{\theta . V_{OD} << 1} I_D \approx \frac{K}{2}(V_{OD}{}^2 - \theta . V_{OD}{}^3) : \text{No longer a square-law device!}$$

Blocker or jammer





- Consider the input spectrum of a weak desired signal and a "blocker"

$$v_i = \underbrace{V_b \cos \omega_1 t}_{Blocker} + \underbrace{V_d \cos \omega_2 t}_{desired}$$

- A strong blocker reduces system gain. This "desensitization" can occur even if the interferer is at a very different frequency.
- This interfering signal is also often called a *jammer*.

Blocking

- The linear terms do not induce any desensitization.
- Second-order terms only generate second harmonic, intermodulation, and DC terms, but no fundamental.
- A cubic nonlinearity ($c_3 v_i^3$), on the other hand, *does* cause desensitization:

$$v_i^3 = V_b^3 \cos^3 \omega_1 t + V_d^3 \cos^3 \omega_2 t + 3V_d^2 V_b \underbrace{\cos^2 \omega_2 t}_{\frac{1}{2}(1+2\cos 2\omega_2 t)} \cos \omega_1 t + 3V_b^2 V_d \underbrace{\cos^2 \omega_1 t}_{\frac{1}{2}(1+2\cos 2\omega_1 t)} \cos \omega_2 t$$

- The first two terms are at the fundamental and third harmonic.
- The last two terms generate terms at ω_1 and ω_2, as well as the intermodulation terms.
  - The third term is much smaller than the last ($V_d << V_b$).

Blocking (cont.)

- Blocker action, therefore, is explained by the last term:

$$\frac{3}{2}c_3 V_b^{\,2} V_d \cos\omega_2 t$$

- In most systems $c_3/c_1 < 0$ (compressive nonlinearity) and therefore the blocker term subtracts from the desired signal. Therefore:

$$\text{Apparent gain} = \frac{c_1 V_d + \frac{3}{2}c_3 V_b^{\,2} V_d}{V_d} = c_1\left(1 + \frac{3}{2}\frac{c_3}{c_1}V_b^{\,2}\right)$$

Out-of-band 3dB desensitization

- Now we can find the blocker level that can desensitize, or simply *desense*, the amplifier by 3dB.

$$20\log(1+\frac{3}{2}\frac{c_3}{c_1}V_b^2) = -3dB$$

- Therefore the power of a 3dB-desensing blocker is given by:

$$P_{OB} = P1\mathrm{dB} + 1.2dB$$

How can we improve linearity?

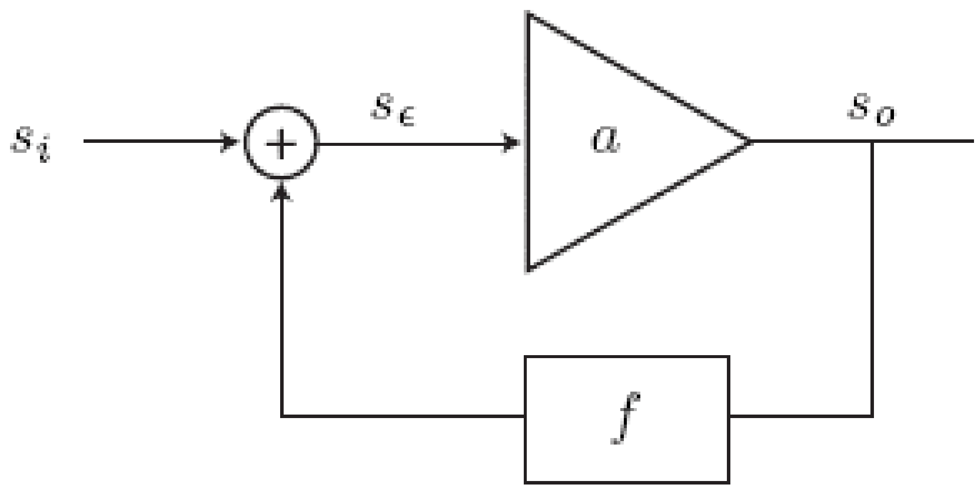
- The idea is to reduce the nonlinear coefficients compared to the linear term.
- For a long-channel MOS device we have

$$I_D = \underbrace{\frac{K}{2}(V_Q - V_{th})^2}_{c_0} + \underbrace{K(V_Q - V_{th})}_{c_1} v_i + \underbrace{\frac{K}{2}}_{c_2} v_i^2$$

- So all we need to do is to increase the overdrive voltage.
- We can also use feedback. How?

Negative feedback review

- Negative feedback depends on large |*af*| (and a linear *f*) to provide significant benefit.



Negative feedback review (cont.)

- If we express the output as:

$$s_o = b_1 s_i + b_2 s_i^2 + b_3 s_i^3 + \ldots$$

- Then the first-order term is:

$$b_1 = \frac{a_1}{1 + a_1 f} = \frac{a_1}{1 + T}$$

Negative feedback review (cont.)

- If we solve for the second-order term we get:

$$b_2 = \frac{a_2}{(1+a_1 f)^3} = \frac{a_2}{(1+T)^3}$$

- And solving for the third-order term yields:

$$b_3 = \frac{a_3(1+T) - 2a_2^2 f}{(1+T)^5}$$

  - Recall that b_3 can be made zero if $a_3(1+T) = 2a_2^2 f$

How feedback changes HD2 and HD3

- From our previous derivation we have

$$HD2 = \frac{1}{2}\frac{b_2}{b_1}V_i = \xrightarrow{V_o=\frac{a_1}{1+T}V_i} \boxed{\underbrace{\frac{1}{2}\frac{a_2}{a_1^2}V_o}_{\text{w/o feedback}}\frac{1}{(1+T)}}$$

- ***For a given output level*** *HD*2 thus improves by (1+*T*).

$$HD3 = \frac{b_3}{4b_1}V_i^2 = \boxed{\underbrace{\frac{a_3}{4a_1^2}V_o^2}_{\text{w/o feedback}}\frac{1}{(1+T)}[1-\frac{2a_2^2 f}{a_3(1+T)}]}$$

- Note that *HD*3 can be forced to zero and also scales by (1+*T*), again **for a given output power**.

Loop transmission effects

- These derivations show that – for equal output levels – the fractional harmonic distortion is inversely proportional to the loop transmission magnitude.
  - The larger the loop transmission, the better.

- Achieving large loop gains may be difficult at high frequencies or over large bandwidths.

- Maintaining closed-loop stability may be difficult with large loop transmissions over a large bandwidth.

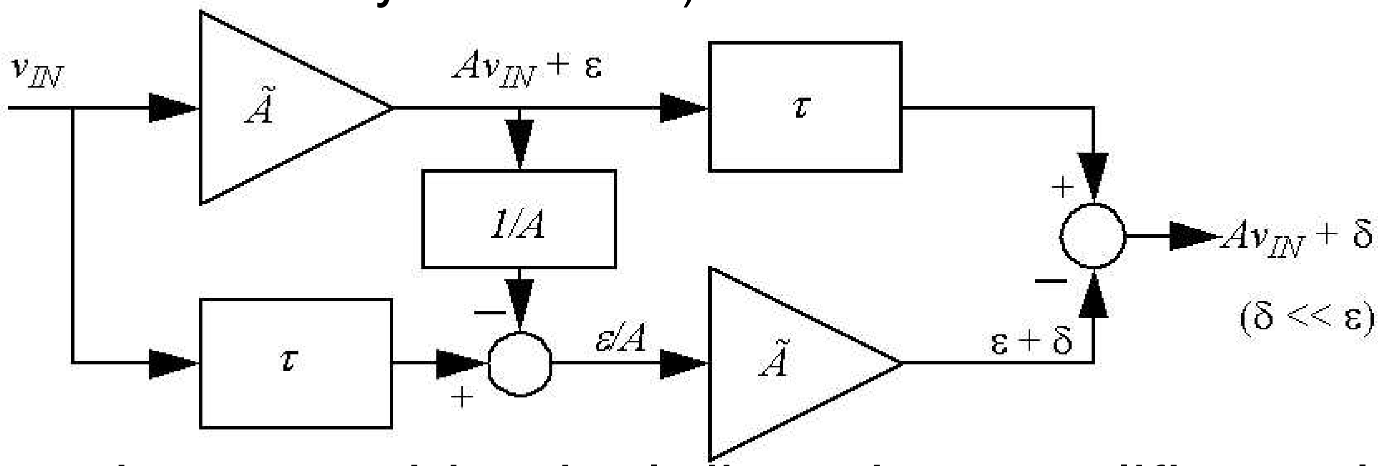
Evading limits: Feedforward

- The need for large loop gains with simple dynamics is not readily satisfied over large bandwidths or at high frequencies.
  - Negative feedback may be impractical in many cases of interest.

- An alternative to negative feedback is feed*forward*.
  - In some sense it involves "less than no feedback."
  - No loop, so bandwidth not constrained by instability.
  - Achievable distortion reduction limited by matching.

Evading limits: Feedforward

- Invented by Harold Black in 1923 (four years before inventing electronic negative feedback).
  - Great idea, but ahead of its time ("matched vacuum tubes" is unintentionally humorous).



- Depends on two identical distorting amplifiers, plus linear attenuator and summing blocks.
  - Also need time delay elements that match amplifier delay for best broadband performance.

Evading limits: Feedforward

- Describe each amplifier by a power series:

$$v_{out} = Av_{in} + \sum_{n=2}^{\infty} c_n v_{in}^n = Av_{in} + \varepsilon$$

- The total "error" (distortion) at the output of the first amplifier is then

$$\varepsilon = \sum_{n=2}^{\infty} c_n v_{in}^n$$

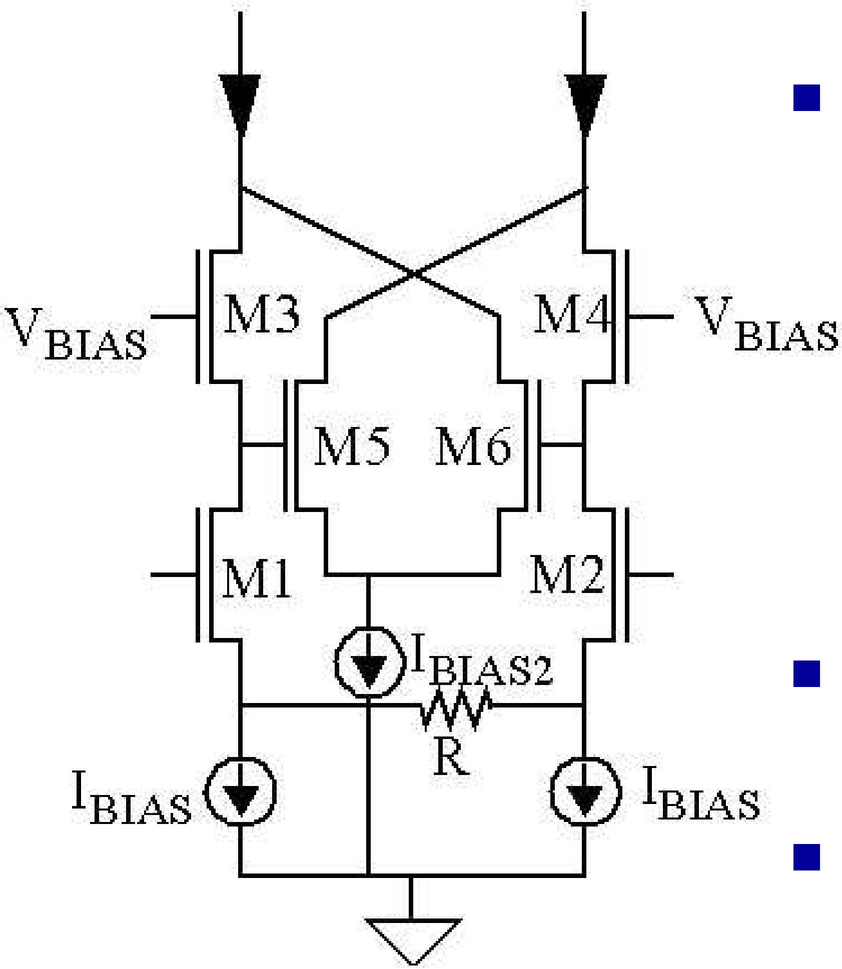
- The total error at the output of the complete amplifier is

$$\delta = \sum_{n=2}^{\infty} \frac{c_n}{A^n} \left[\sum_{n=2}^{\infty} c_n v_{in}^n \right]^n$$

Evading limits: Feedforward

- Note the attenuation by A^n in the error terms of the final output.
  - High-order terms quickly die off as n increases.
  - Large A helps, too.
- There is no need to satisfy a phase margin constraint, because there is no feedback loop.
  - Distortion reduction can be effective over a broad band.
  - Useful bandwidth depends on how well the delay elements track amplifier delay.

Classic feedforward example: Cascomp



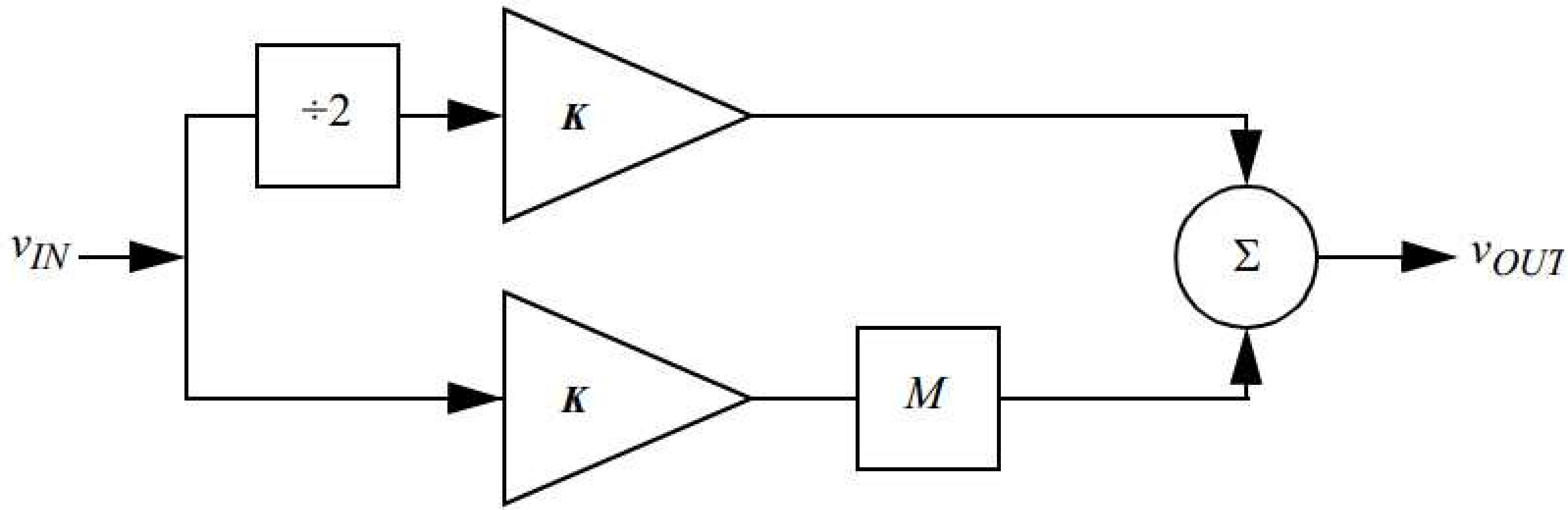
- M1-M4 are a differential cascode with emitter degeneration.
  - Perfect transconductance would put all of v_{in} across resistor.
  - Error is due to change in v_{gs1}, v_{gs2} (= v_{gs3}, v_{gs4}).
- M5-M6 measure the error, invert it, and feed it forward.
- Invented by Pat Quinn, used in Tektronix 2465 'scope.

“Weak” or “soft” feedforward

- Many variations on the feedforward idea.
- Black’s architecture aims to reduce all distortion terms.
  - *Strong feedforward* sometimes used to describe it.
- *Weak* or *soft* feedforward targets a finite subset of distortion terms.
  - Common choice is to suppress second- or third-order distortion product.

“Weak” or “soft” feedforward

- Here’s a block-diagram representation of a weak feedforward system:



- What value of *M* (if any) cancels out the third-order distortion term?

When good circuits go bad

Why RF design is hard

- Can’t ignore parasitics:
  - 100fF is 320Ω@5GHz; 1.6Ω@1THz
- Can’t squander device power gain.
- Can’t tolerate much noise or nonlinearity.
  - 1μV amplitude = 10fW@50Ω
- Can’t expect accurate models, but you still have to ship anyway.

Traditional RF design flow

H7N9

- Wear pointy wizard hat.

- Obtain chicken.
- Design first-pass circuit.
- Utter magical Latin incantations (“*semper ubi sub ubi*...*omnia pizza in octo partes divisa est...e pluribus nihil*”).
- Test circuit. Weep.
- Adjust chicken. Iterate.

© Ron Leishman * www.ClipartOf.com/438685

Dark secrets: A partial list

- MOSFETs: Textbook lies
- The two-port noise model: Why care?
  - Optimum noise figure vs. maximum gain
- Secrets about impedance matching
- Secrets about linearity and time invariance
- Mixers: Deceptions, myths and confusion
- Gain-bandwidth limits aren't
- Strange impedance behaviors (SIBs)

MOSFETs: What Your Textbooks May Not Have Told You

The standard story

- "Gate-source impedance is that of a capacitor."

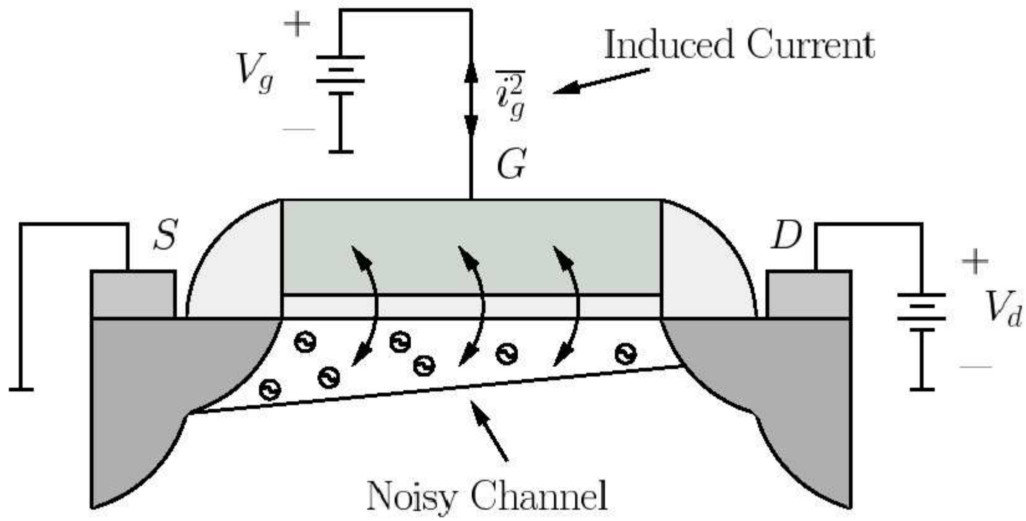
- A capacitor is lossless. A capacitive gate-source impedance therefore implies *infinite* power gain at all frequencies.

Gate impedance has a real part

- Gate-source impedance *cannot be* purely capacitive.
  - True even if gate material is superconductive.
- Phase shift associated with finite carrier transit speed means gate field does nonzero work on channel charge.
- Therefore, power gain is not infinite.
- *There is also noise associated with the dissipation.*

Noisy channel charge

- Fluctuations couple capacitively to both top and bottom gates.
  - Induces noisy gate currents.
  - Bottom-gate term is ignored by most models and textbooks.



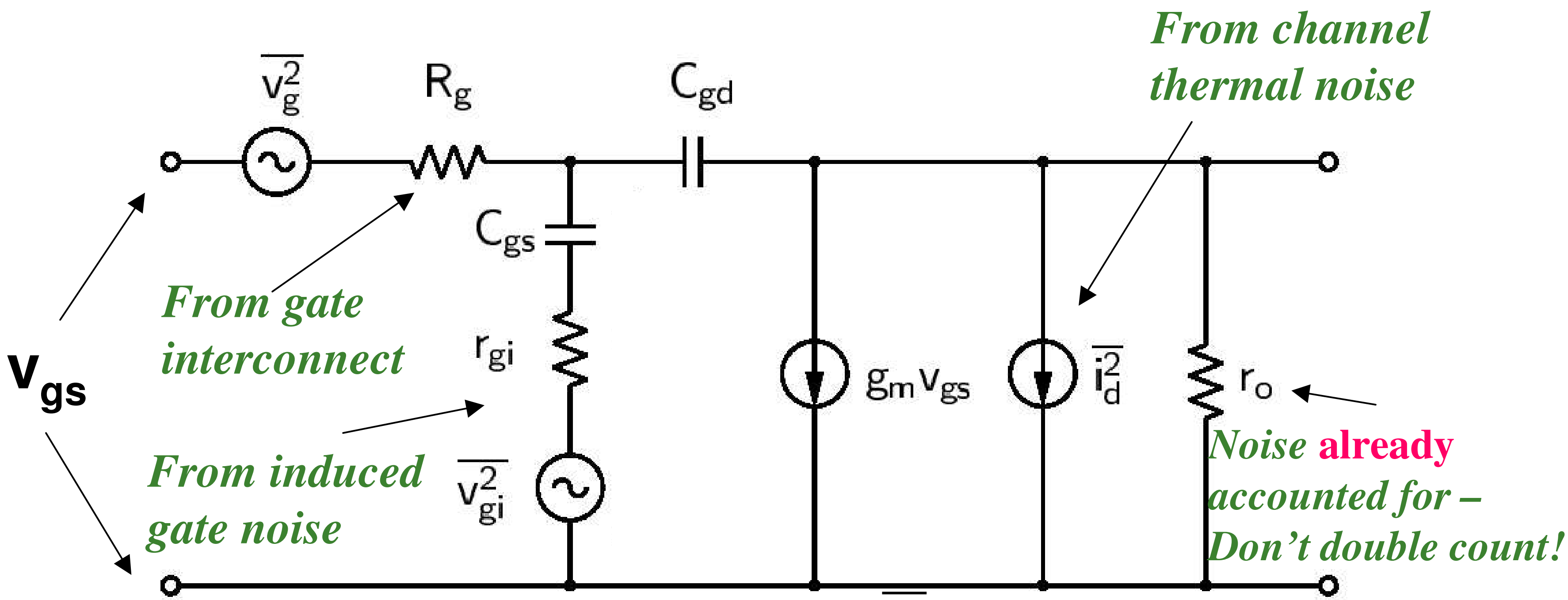
[Shaeffer]

- Gate noise current. $\overline{i_g^2} = 4kTB\delta \frac{(\omega C_{gs})^2}{5g_{d0}}$
- Real component of Y_g. $\mathrm{Re}\left[Y_g\right] = \frac{(\omega C_{gs})^2}{5g_{d0}}$

Sources of noise in MOSFETs

- (Thermally-agitated) channel charge.
  - Produces current noise in both drain *and* gate.
- Interconnect resistance.
  - Series gate resistance R_g is very important.
- Substrate resistance.
  - Substrate thermal noise modulates back gate, augments drain current noise in some frequency range.

(All) FETs and gate noise: Basic model



$$\overline{i_d^2} = 4kT\gamma g_{d0}$$

$$\overline{v_{gi}^2} = 4kTB\delta r_{gi}$$

$$r_{gi} = \frac{1}{5g_{d0}}$$

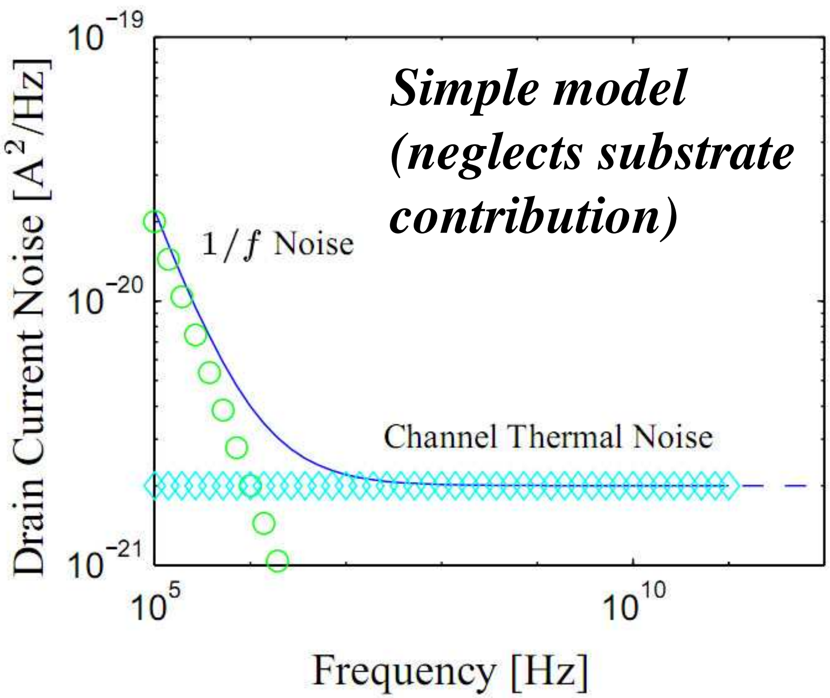
(Note the placement of v_{gs}.)

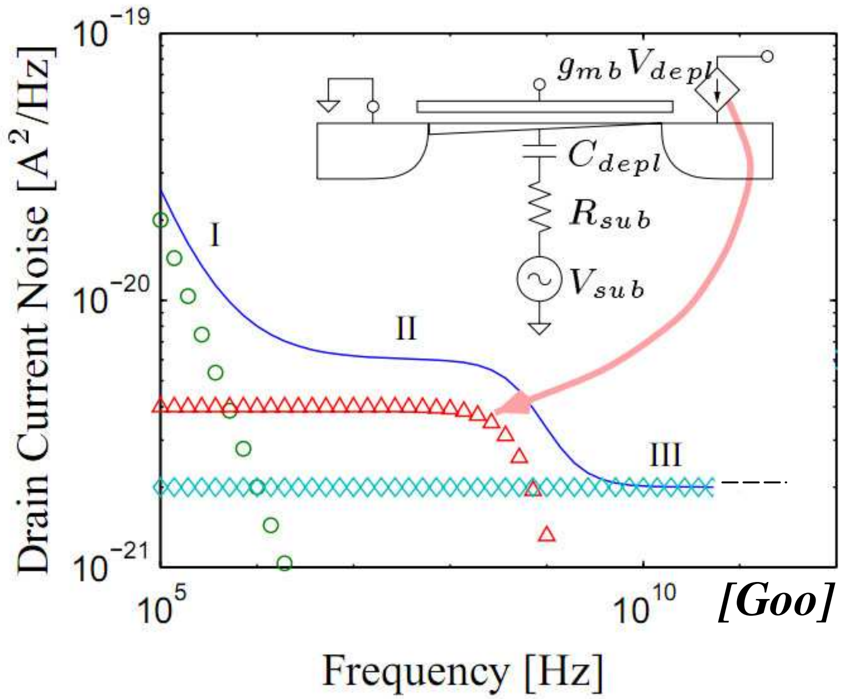
Important: Common error is to define Vgs as across Cgs alone.

Dark secret: Confusing noise models

- Measuring drain noise at different frequencies has led to disagreement about γ.
  - Measurements made below ~1GHz may suggest both “excess” noise (γ) and sensitivity to the number of substrate taps, depending on model used.
- Early fears that deep-submicron MOSFETs suffer from significant increase in γ not borne out.

Substrate noise effects



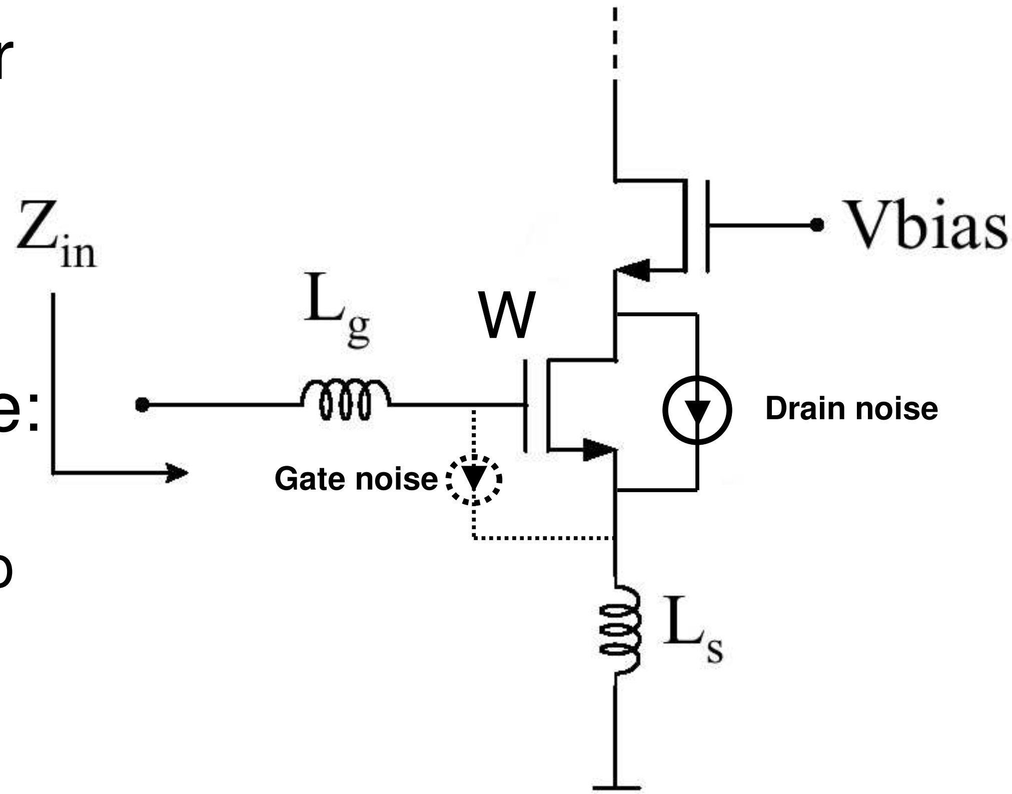


[Goo]

$$S_{i_d,sub} = \frac{4kTR_{sub}g_{mb}^2}{1+(\omega R_{sub}C_{depl})^2}$$

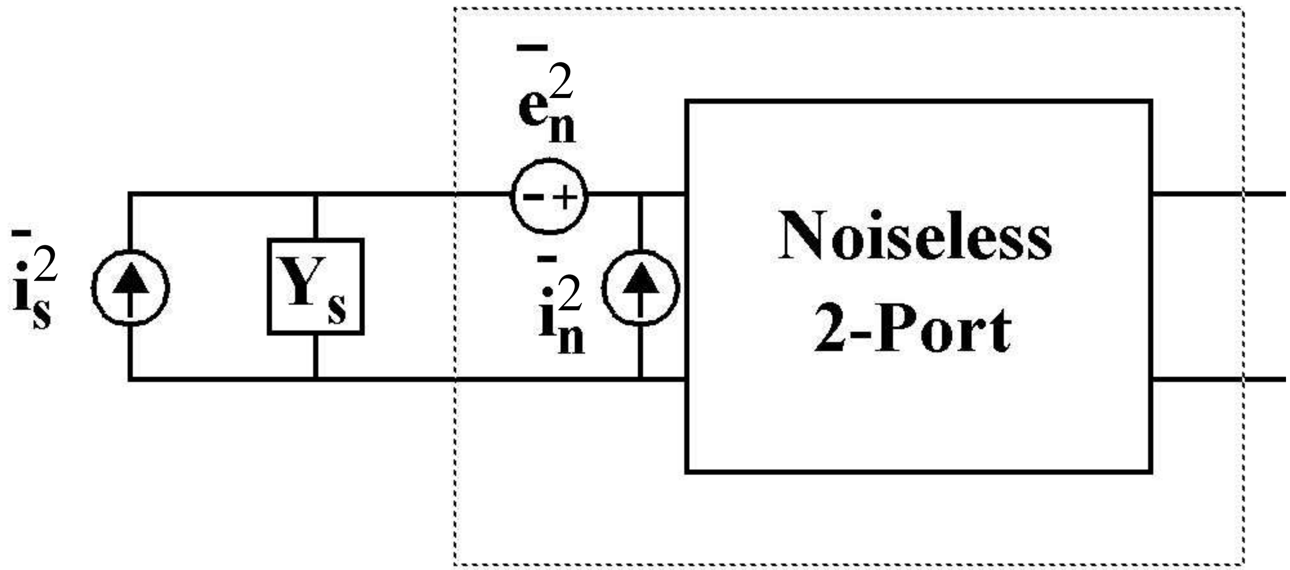
Dark secret: Gate noise is real

- Let W → 0 while maintaining resonance and current density (for fixed f_T).
  - Gain stays fixed.
  - I_{bias} → 0.
- If you ignore gate noise:
  - Output noise → zero; absurd to consume zero power and provide noiseless gain.



The Two-Port Noise Model: Why?

Two-port noise model



$$F = \frac{\overline{i_s^2} + \overline{\left|i_n + Y_s e_n\right|^2}}{\overline{i_s^2}} \quad = \quad 1 + \frac{\overline{i_u^2} + \left|Y_c + Y_s\right|^2 \overline{e_n^2}}{\overline{i_s^2}}$$

- The *IRE* chose not to define *F* directly in terms of equivalent input noise sources. Instead:

Two-port noise model

Let

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f},$$

$$G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f},$$

and

$$G_s \equiv \frac{\overline{i_s^2}}{4kT\Delta f}.$$

Conditions for minimum noise figure

$$B_s = -B_c = B_{opt}$$

$$G_s = \sqrt{\frac{G_u}{R_n} + G_c^2} = G_{opt}$$

$$F_{min} = 1 + 2R_n[G_{opt} + G_c] = 1 + 2R_n\left[\sqrt{\frac{G_u}{R_n} + G_c^2} + G_c\right].$$

$$F = F_{min} + \frac{R_n}{G_s}\left[(G_s - G_{opt})^2 + (B_s - B_{opt})^2\right]$$

Dark secret: Murphy wants to hurt you

- Minimum NF and maximum power gain occur for the same source Z *only if three miracles occur simultaneously*:
  - $G_c = 0$ (noise current has no component in phase with noise voltage); *and*
  - $G_u = G_n$ (conductance representing uncorrelated current noise equals the fictitious conductance that produces noise voltage); *and*
  - $B_c = B_{in}$ (correlation susceptance happens to equal the actual input susceptance)

Dark secret: You don't always want to match impedances

Why match?

- Conjugate match maximizes power transfer.
- Terminating a *T*-line in its characteristic impedance makes the input impedance length-independent.
  - Also minimizes peak voltage and current along line.
- Choosing a standard impedance value (e.g., 50Ω) facilitates assembly, fixturing and instrumentation; it addresses *macroscopic* concerns.

Why not match?

- Amplifiers generally exhibit best noise figure with a non-optimal gain match.
- Many amplifiers are more stable or robust (in the PVT sense) when appropriately mismatched.
- If power gain is abundant (and stability and noise are not a problem), can afford mismatch, resulting in a simpler circuit.

Dark secret: Circuits are neither linear nor time-invariant

LTI, LTV and all that

- A system is linear if superposition holds.
- A system is TI if an input timing shift only causes an equal output timing shift.
  - Shapes stay constant.
- If a system is LTI, it can only scale and phase-shift Fourier components.
  - Output and input frequencies are the same.
- If a system is LTV, input and output frequencies can be different, ***despite being linear***.
- If a system is nonlinear, input and output frequencies will generally differ.

Mixers are supposed to be linear!

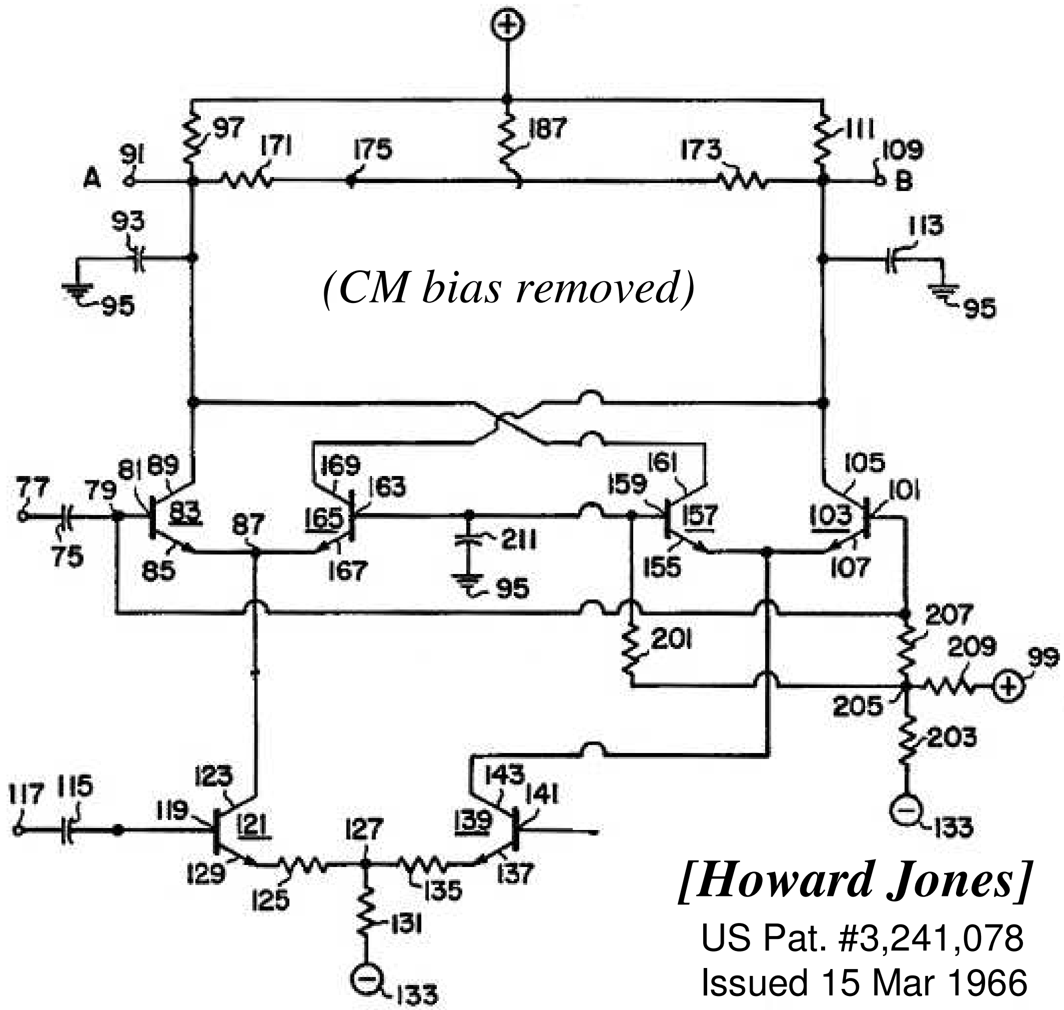
- But they are *time-varying* blocks.
  - Too many textbooks and papers say “mixers are nonlinear…” Mixers are nonlinear in the same way that amplifiers are nonlinear: *Undesirably*.

- Mixers are noisier than LNAs for reasons that will be explained shortly. NF values of 10-15dB are not unusual.

- Main function of an LNA is usually to provide enough gain to overcome mixer noise.

Dark secret: Most "Gilbert" mixers aren't

- This is a *Jones* mixer.
  - Most textbooks and papers (still) wrongly call it a Gilbert cell.
- A true Gilbert cell is a *current-domain* circuit, and uses *predistortion* for linearity.



[Howard Jones]

US Pat. #3,241,078
Issued 15 Mar 1966

The mixer: An LTV element

- Whether Gilbert, Jones or Smith, modern mixers depend on *commutation* of currents or voltages.

- We idealize mixing as the equivalent of multiplying the RF signal by a square-wave LO.
  - Single-balanced mixer: RF signal is unipolar.
  - Double-balanced mixer: RF signal is DC-free.

- Mixing is ideally *linear*: Doubling the input (RF) voltage should double the output (IF) voltage.

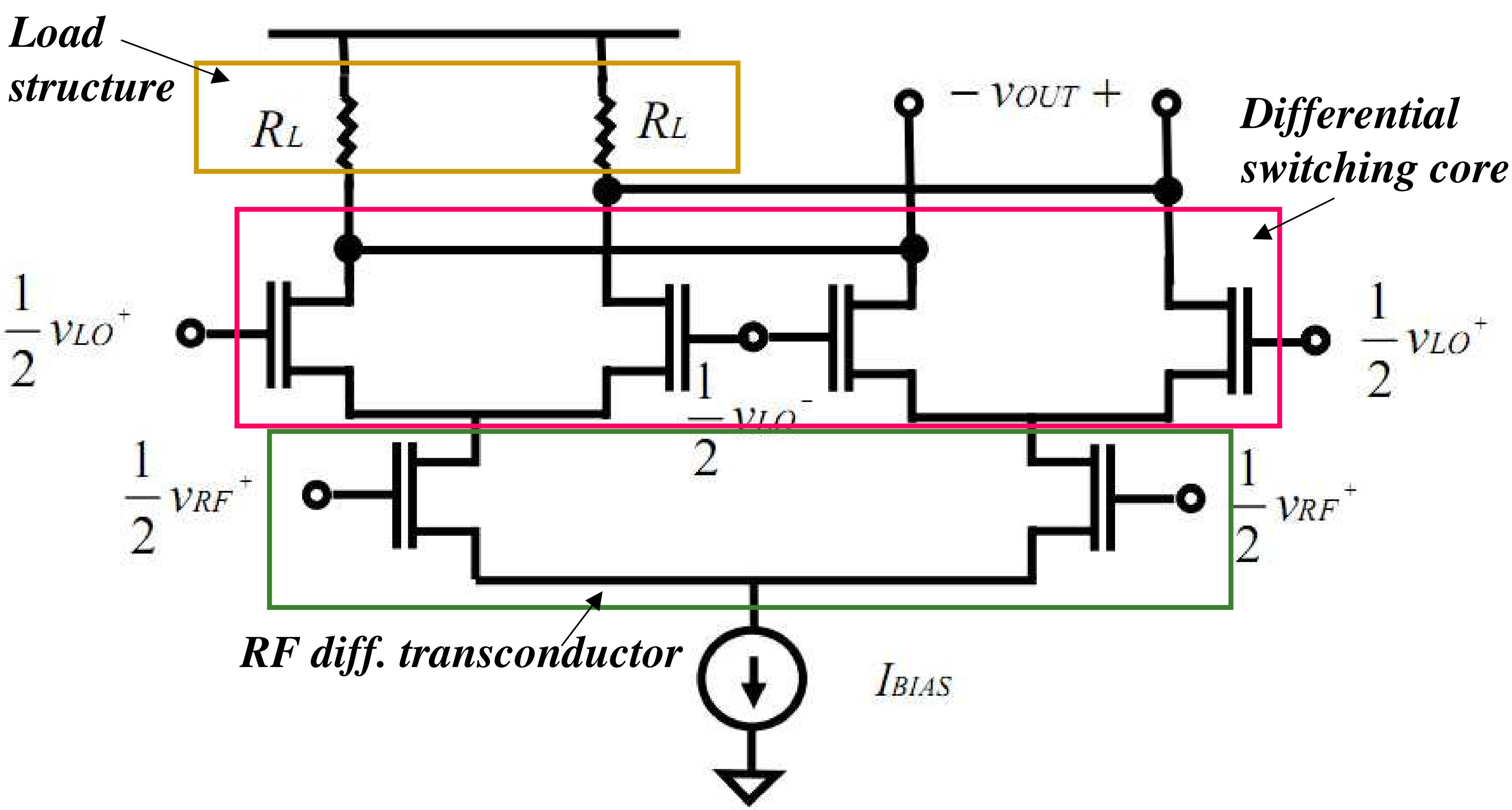
A multiplier is an ideal mixer

- Key relationship is:

$$\boxed{A\cos\omega_1 t}\cos\omega_2 t = \frac{A}{2}[\cos(\omega_1+\omega_2)t + \cos(\omega_1-\omega_2)t]$$

- Can be thought of as an amplifier with a time-varying amplification factor (e.g., term in blue box, above).

Sources of noise in mixers



Mixer noise

- Load structure is at the output, so its noise adds to the output directly; it undergoes no frequency translations.
  - If $1/f$ noise is a concern, use PMOS transistors or poly resistor loads.

- Transconductor noise appears at same port as input RF signal, so it translates in frequency the same way as the RF input.

Dark secret: Switching noise can *dominate*

- Instantaneous switching not possible.
  - Noise from switching core passing through linear region can actually *dominate*.
  - Common-mode capacitance at tail nodes of core can reduce effectiveness of large LO amplitudes.

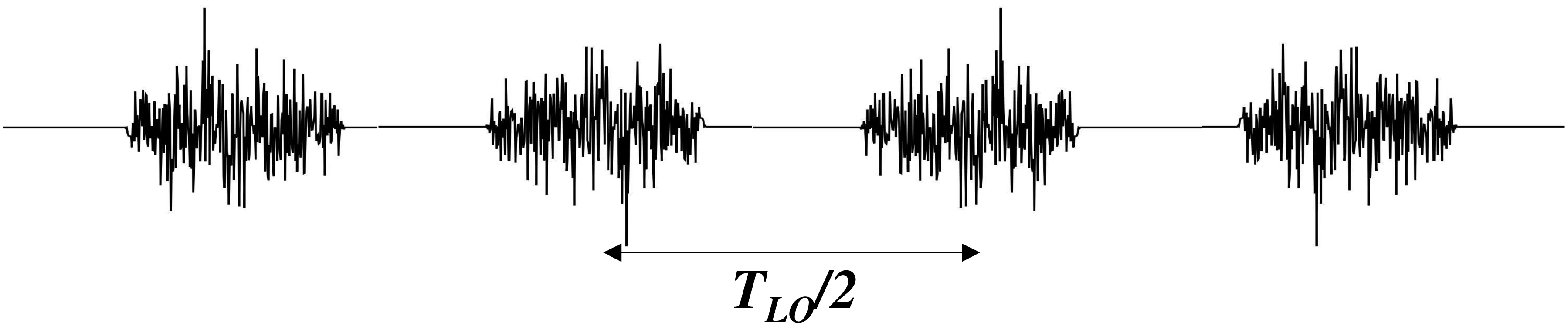
- Periodic core switching is equivalent to windowed sampling of core noise at (twice) the LO rate.
  - Frequency translations occur due to this self-mixing.

Noise contribution of switching core

- As switching transistors are driven through the switching instant, they act as a differential pair for a brief window of time t_s.
  - During this interval, the switching transistors transfer their drain noise to the output.
  - Changing drain current implies a changing PSD for the noise; it is cyclostationary.

Noise contribution of switching core

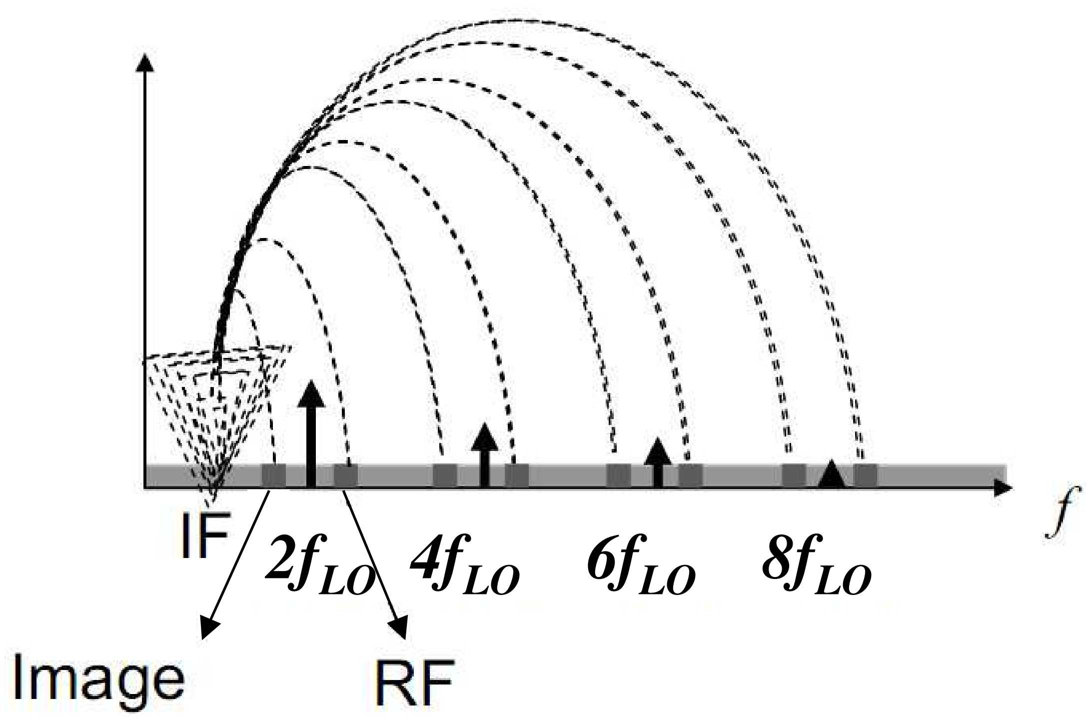
- The noise contributed by the switching core appears roughly as follows:



- Mathematically equivalent to multiplying stationary noise by a shaped pulse train of fundamental frequency $2f_{LO}$.

Noise contribution of switching core

- Noise at $2nf_{LO}$ +/- f_{IF} will therefore translate to the IF. This noise folding partly explains the relatively poor noise figure of mixers.



Terrovitis mixer noise figure equation

- A simplified analytical approximation for the SSB noise figure of a Jones mixer is

important

$$F_{SSB} \approx \frac{\alpha}{c^2} + \frac{2\gamma g_m \alpha + 4\gamma \overline{G} + G_L}{c^2 g_m^2 R_S}$$

- Here, g_m is the transconductance of the bottom differential pair; G_L is the conductance of the load; R_S is the source resistance, and γ is the familiar drain noise parameter.
  - See *[Terrovitis]* for more complete version.

Terrovitis mixer noise figure equation

- The parameter $\overline{G}$ is the time-averaged transconductance of each pair of switching transistors. For a plain-vanilla Jones mixer,

$$\overline{G} \approx \frac{2I_{BIAS}}{\pi V_{LO}}$$

- The parameter α is related to the sampling aperture, and has an approximate value

$$\alpha \approx 1 - \frac{4}{3} t_s f_{LO}$$

Terrovitis mixer noise figure equation

- The parameter c is directly related to the effective aperture, and is given by

$$c \approx \frac{2}{\pi}\left[\frac{\sin(\pi t_s f_{LO})}{\pi t_s f_{LO}}\right]$$

- This parameter asymptotically approaches $2/\pi$ in the limit of infinitely fast switching.

Extreme Circuit Design:
Amplifier Tricks

Thomas Lee, 17 August 2014

The myth of G-BW limits

- Most circuit design textbooks convey the belief that there is a “gain-bandwidth limit.”
  - Given certain assumptions, yes, G-BW is limited, but that limit isn’t fixed…so it’s not quite as hard a limit.
- Let’s see what happens if we violate those “certain assumptions.”

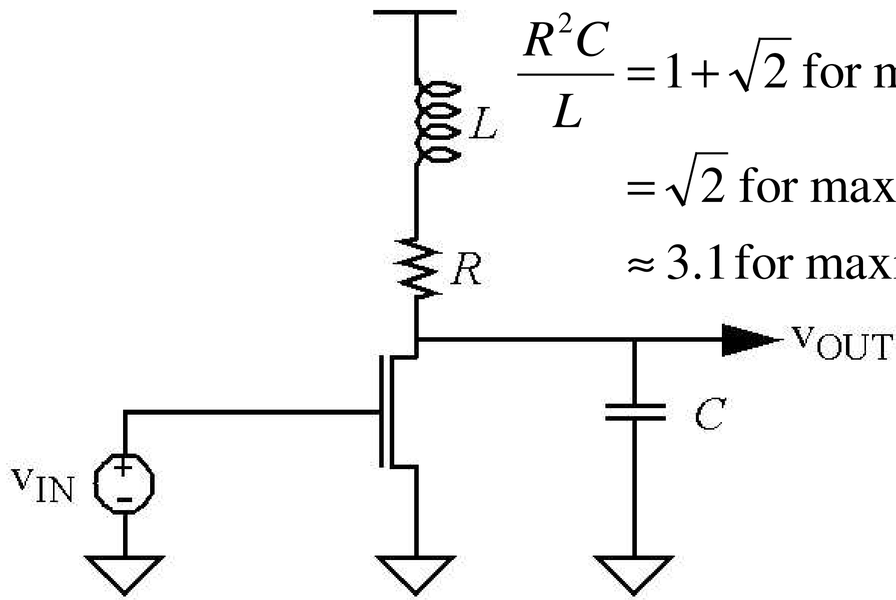
Single-pole common-source

- *Everyone* knows that the gain is $g_m R_L$, -3dB bandwidth is $1/R_L C_L$, so GBW is just g_m/C_L.
  - Single-pole systems trade gain for bandwidth directly.

- But what about higher-order systems? Can we use the additional degrees of freedom to do better?

Shunt peaking

- Ancient idea – dates back to the 1930s.



$$\frac{R^2C}{L} = 1+\sqrt{2} \text{ for maximally flat amplitude}$$

$$= \sqrt{2} \text{ for maximum BW}$$

$$\approx 3.1 \text{ for maximally flat delay}$$

- Adding one low-*Q* inductor nearly doubles BW.
  - 1.85x max. boost; 1.72x boost @ max. flat amplitude, 1.6 boost for max. flat delay.

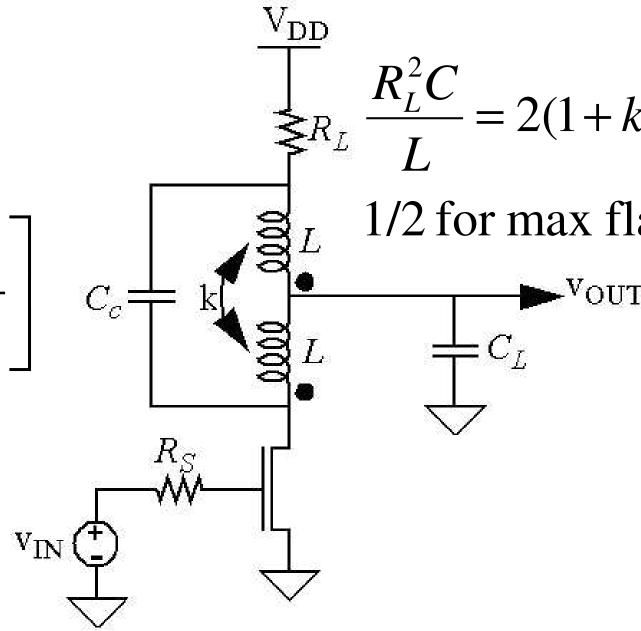
Bridged T-coil

- Used extensively in Tektronix ‘scopes.

$$\frac{R_L^2 C}{L} = 2(1+k);\ \mathrm{k} = 1/3 \text{ for max. flat gain,}$$

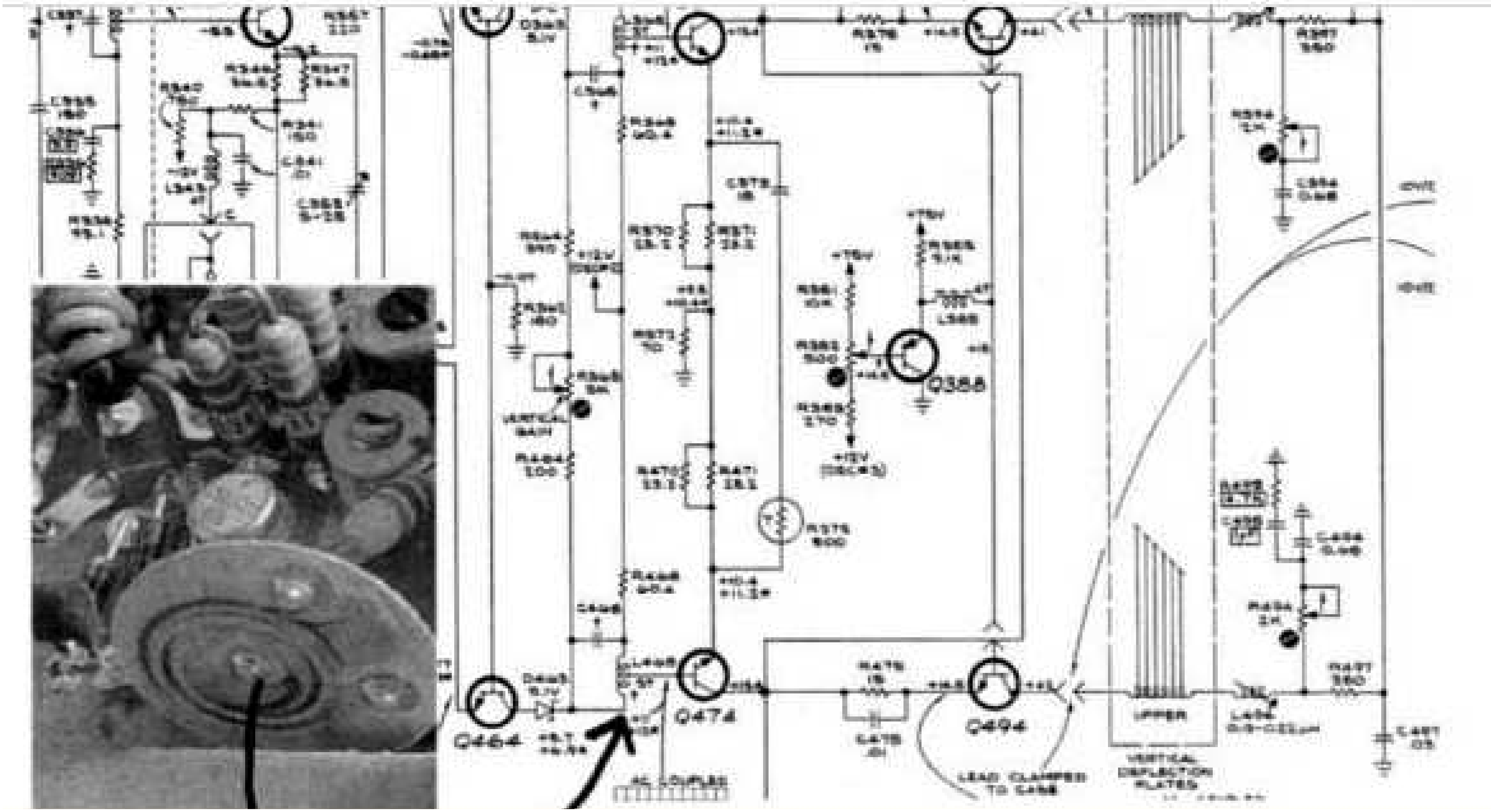
1/2 for max flat delay

$$C_C = \frac{C_L}{4}\left[\frac{1-k}{1+k}\right]$$



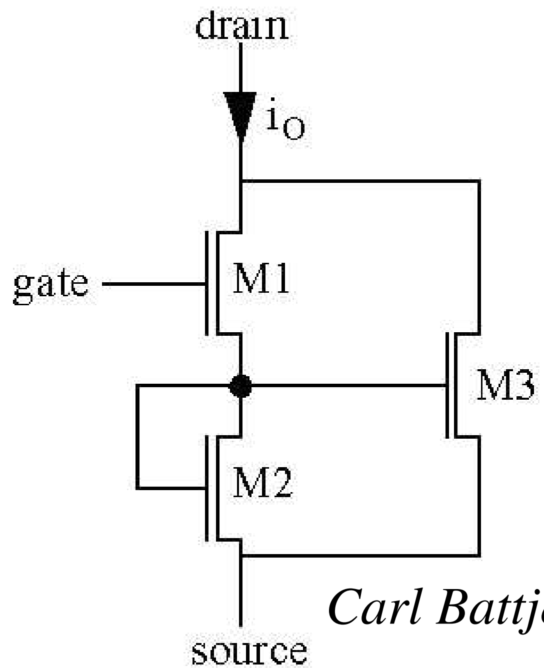
- A small capacitor and low-*Q* transformer nearly triples BW.
  - ~3x max boost; 2.83x boost @ max. flat delay.

Bridged T-coils in Tek 454



1967; last, fastest (150MHz) Tek scope using all discretes

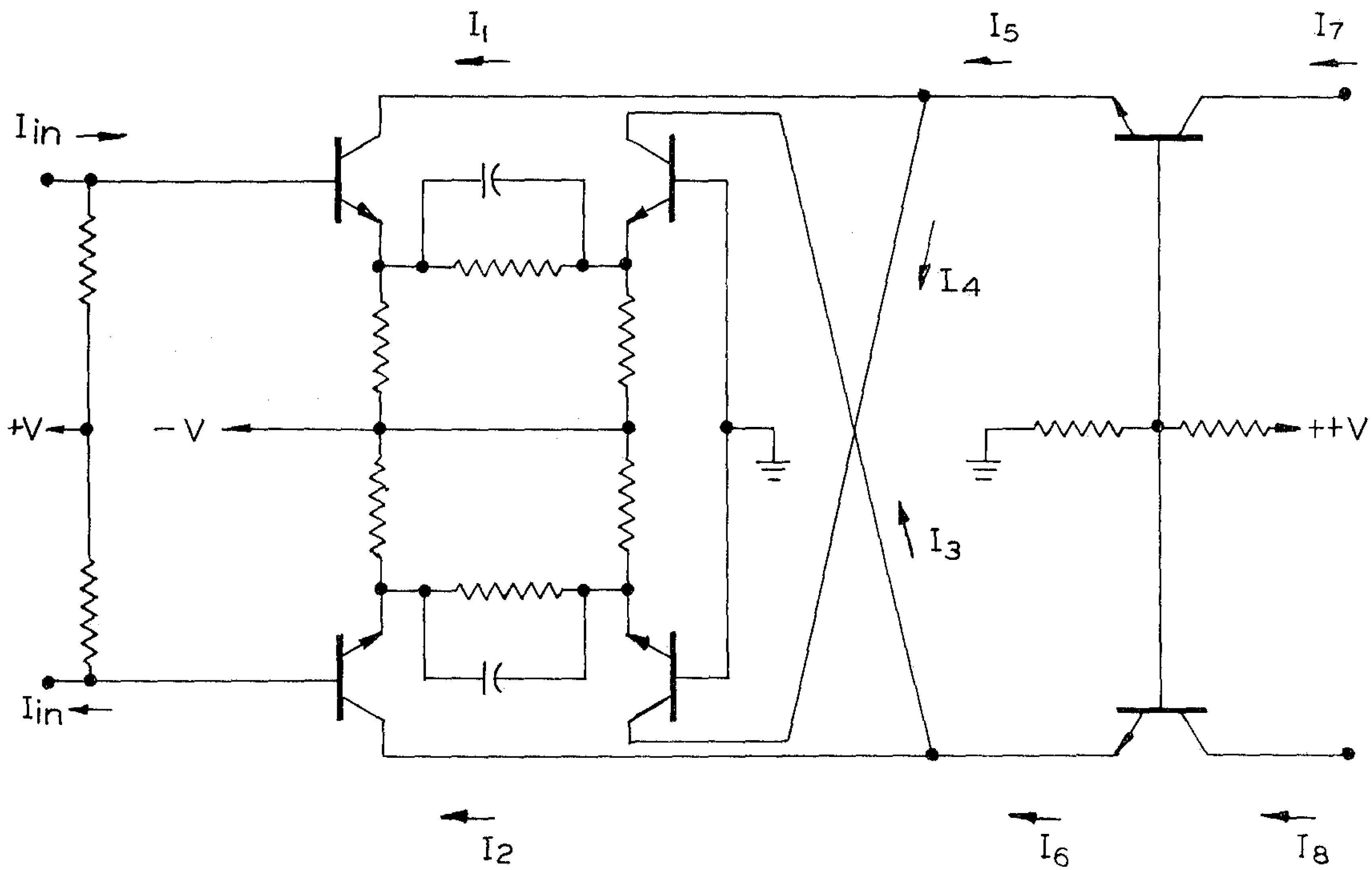
f_T isn't a process limit



Carl Battjes, Tektronix, US Pat. 4236119

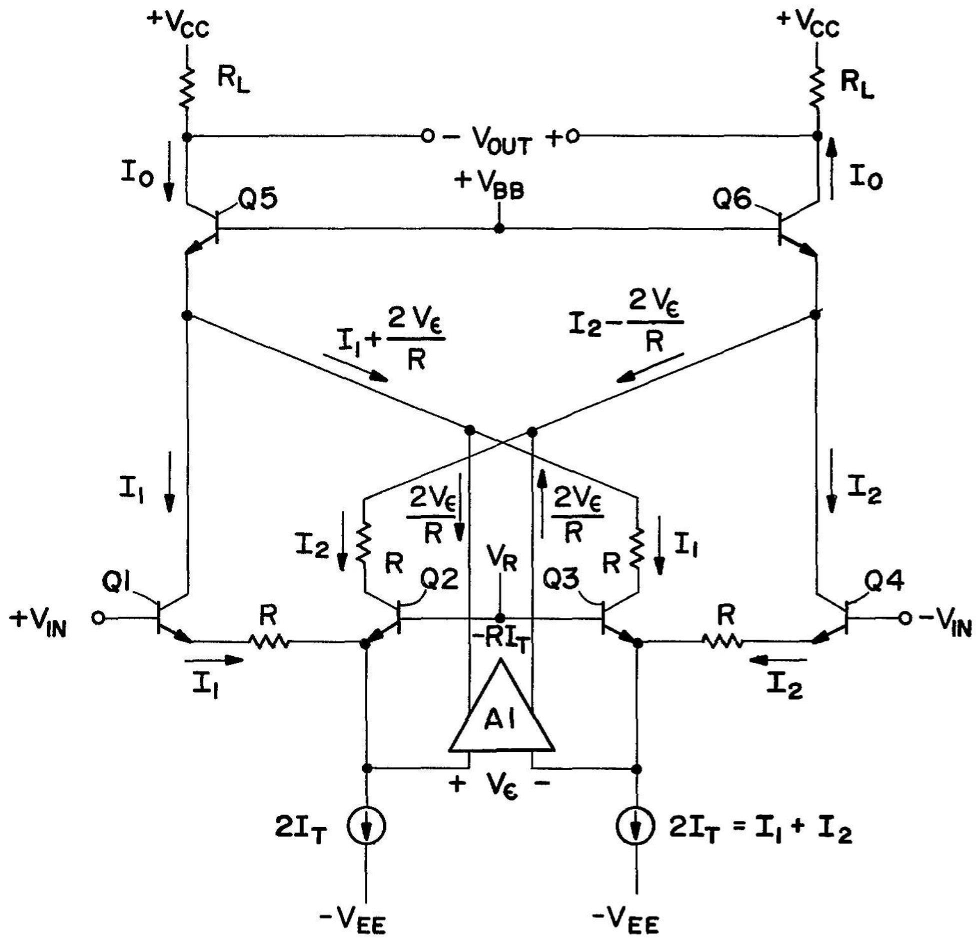
- $\omega_T = g_m/C_{in}$, to an approximation. The Battjes doubler reduces C_{in} to provide a 1.5x boost in f_T. Used with T-coils in Tek 7904 scope to get 500MHz system BW using 3GHz transistors.

Differential, cascoded f_T doubler



Carl Battjes, Tektronix, US Pat. 3633120

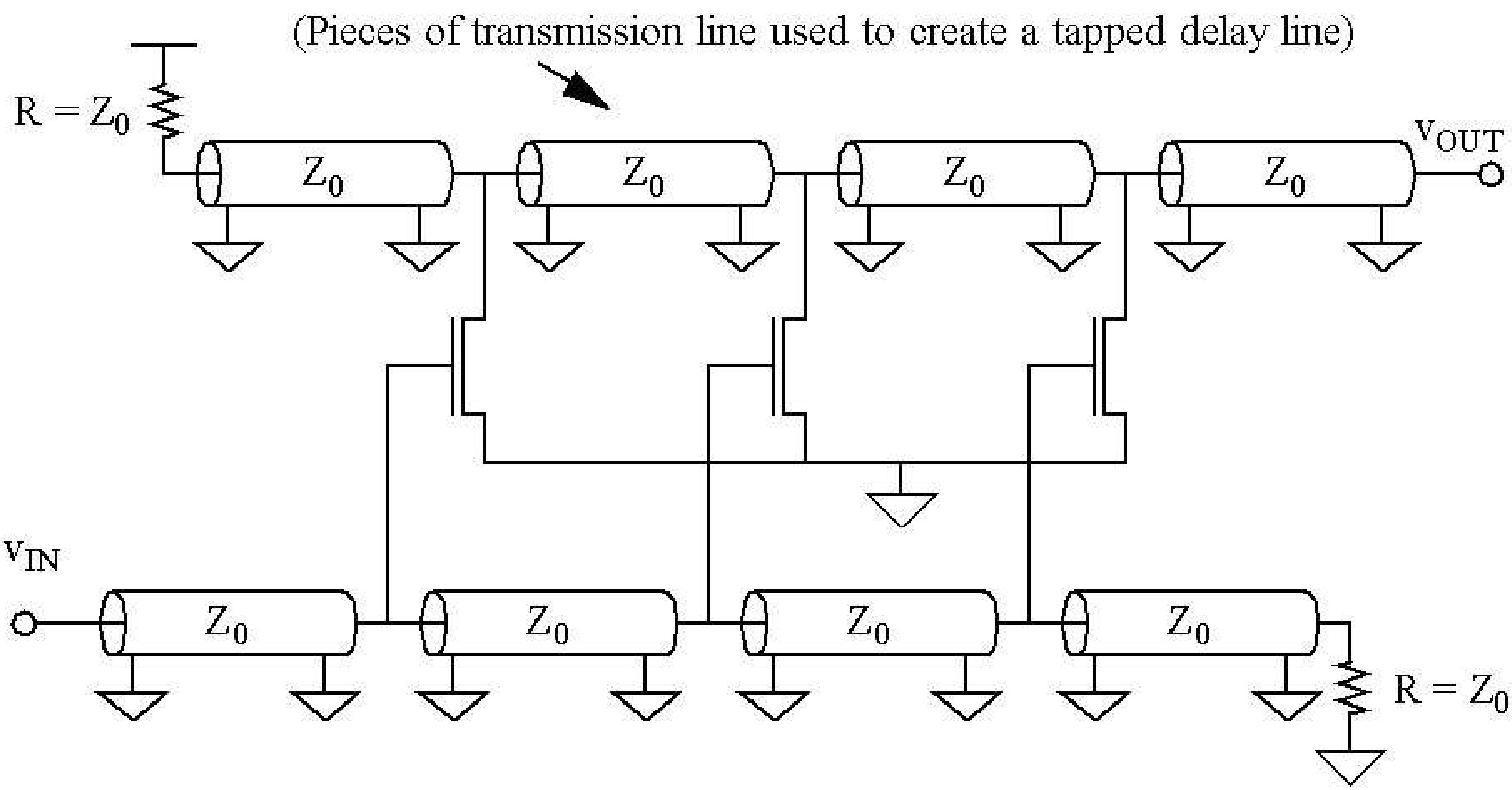
Diff., cascoded doubler w/thermal comp.



Einar Traa, Tektronix, US Pat. 4267516

The distributed amplifier

- Used extensively in Tektronix ‘scopes.



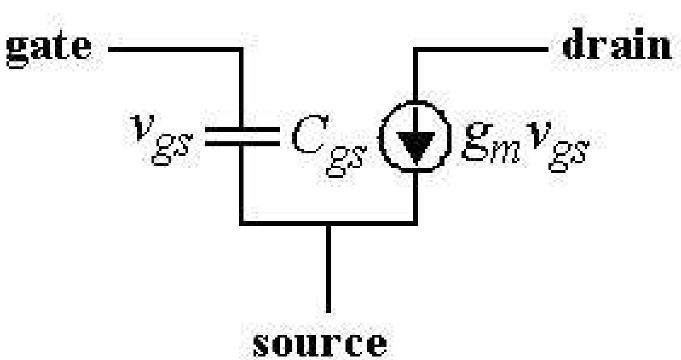
The message?

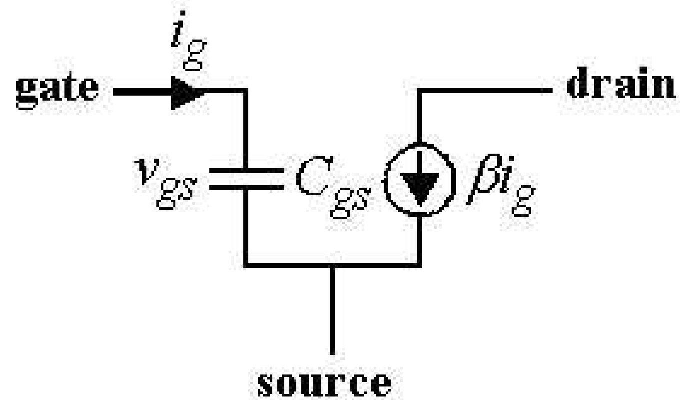
- More poles = more degrees of freedom = greater ability to effect desirable tradeoffs.
- Extend gain-bandwidth tradeoff to gain-bandwidth-*delay* tradeoff.
  - Delay is frequently unimportant, so trading it for GBW is possible in many practical cases.
  - $G\text{-}BW\text{-}T_D$ tradeoff possible *only* if you can create excess delay – you can't trade what you don't have. In turn, creating delay over large BW implies many poles.

Extreme Circuit Design: *Strange Impedance Behaviors (SIBs)*

First: Some simple transistor models

- Can use *either* gate-source voltage *or* gate current as independent control variable



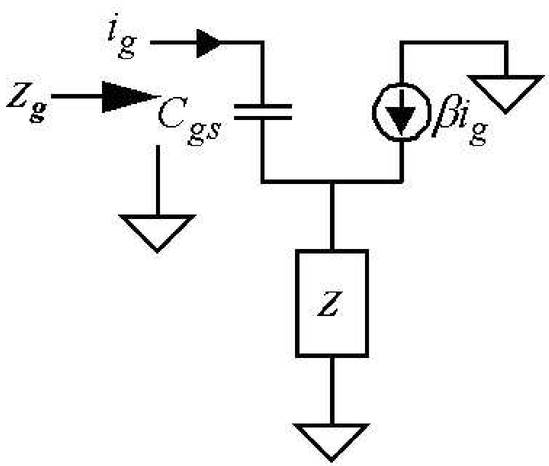


- Models are fully equivalent as long as we choose

$$\beta = \frac{g_m v_{gs}}{i_g} = \frac{g_m}{sC_{gs}} = \frac{\omega_T}{j\omega} = -j\frac{\omega_T}{\omega}$$

View from the gate: Load in source

- Consider input impedance of the following at $\omega \ll \omega_T$:



$$Z_g = \frac{1}{j\omega C_{gs}} + (\beta + 1)Z \approx \frac{1}{j\omega C_{gs}} + (-j\frac{\omega_T}{\omega})Z$$

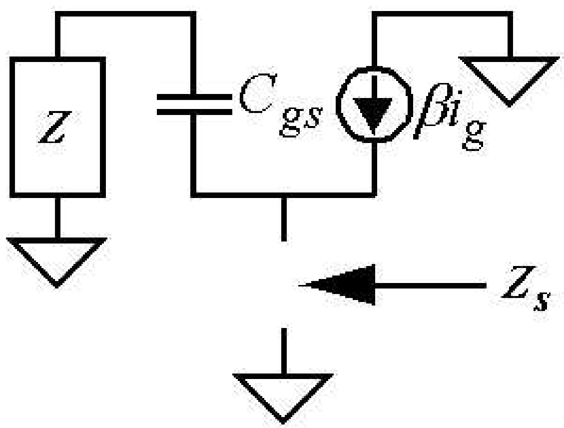
- The non-intuitive behavior comes from the second term: The impedance *Z* gets multiplied by a (negative) imaginary constant.

What does multiplication by $-j\omega_T/\omega$ do?

- Turns R into capacitance = $1/\omega_T R$.
- Turns L into resistance = $\omega_T L = g_m(L/C_{gs})$.
- Turns C into *negative* resistance = $-\omega_T/\omega^2 C$.

View from the source: Load in gate

- Now consider input impedance of the following:



$$Z_s = \frac{\dfrac{1}{j\omega C_{gs}} + Z}{\beta + 1} \approx \frac{1}{g_m} + (j\frac{\omega}{\omega_T})Z$$

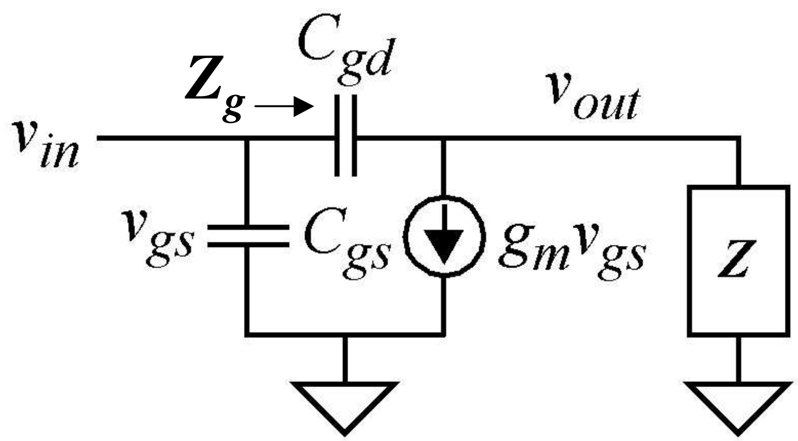
- This time, *Z* gets multiplied by a *+j* factor.

What does multiplication by $+j\omega/\omega_T$ do?

- Turns R into inductance = R/ω_T.
- Turns C into resistance = $1/\omega_T C = (C_{gs}/C)\ (1/g_m)$.
- Turns L into negative resistance = $-\omega^2 L/\omega_T$.

View from the gate: Load in drain

- Consider (partial) input impedance of the following:



$$Z_g \approx \frac{1}{sC_{gd}(1+g_m Z)}$$

$$Z_g \approx \frac{1}{sC_{gd}g_m Z} \quad \text{if } |g_m Z| >> 1$$

- This is the generalized Miller effect: C_{gd} is multiplied by *complex* gain when viewed by gate.

What does multiplication by $g_m Z$ do?

- Turns R into capacitance = $C_{gd}(g_m R)$; this is just the classic Miller effect.

- Turns C into resistance = $(C/C_{gd})(1/g_m)$.

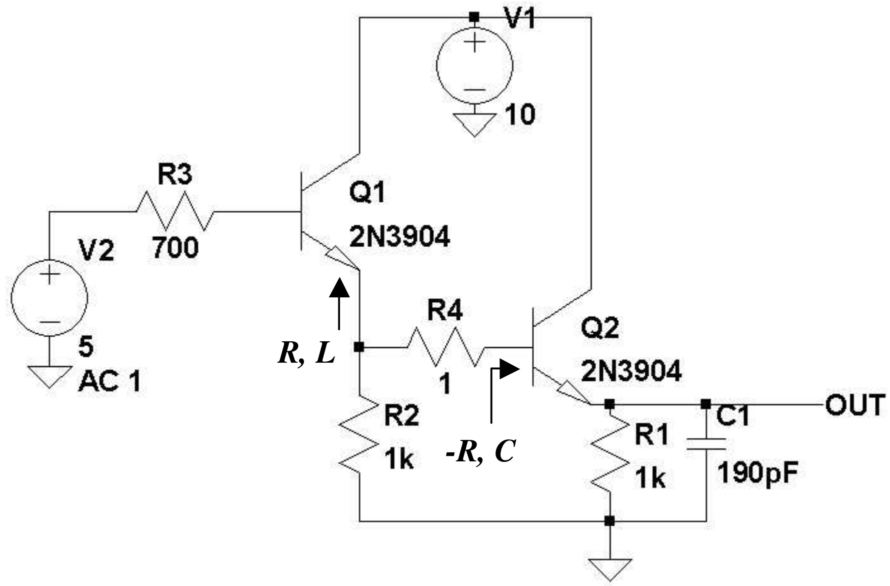
- Turns L into negative conductance = $-\omega^2 g_m L C_{gd}$.

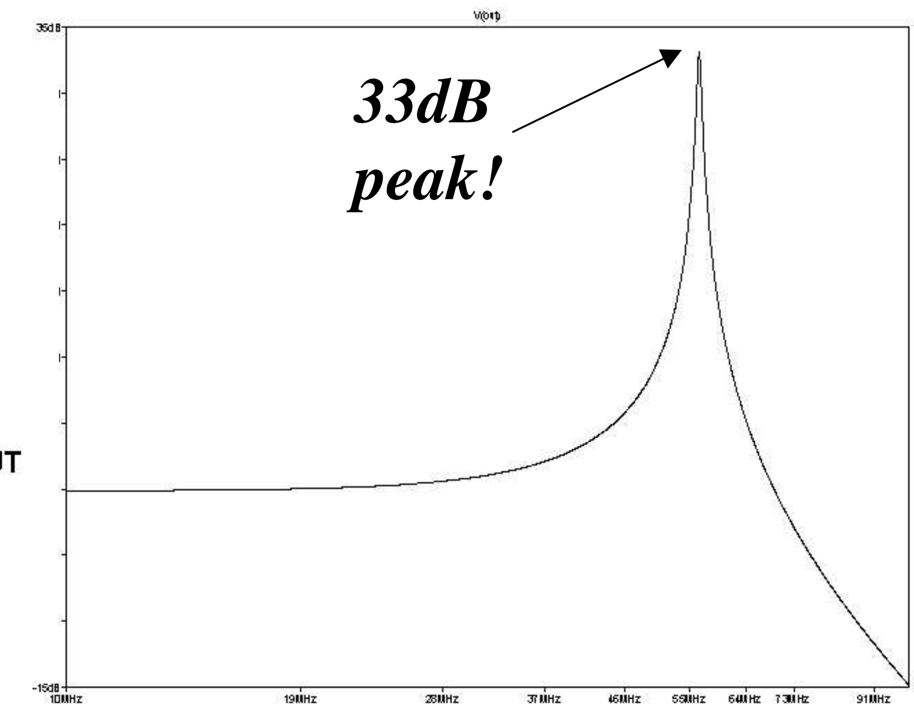
Why SIBs are strange

- Apparent weirdness arises because of feedback around complex gains.
- Phase shift associated with complex gains causes impedances to change *character*, not just magnitude.
- The strangeness evaporates once you spend a little time studying where it comes from.

SIBs example: Cascaded followers

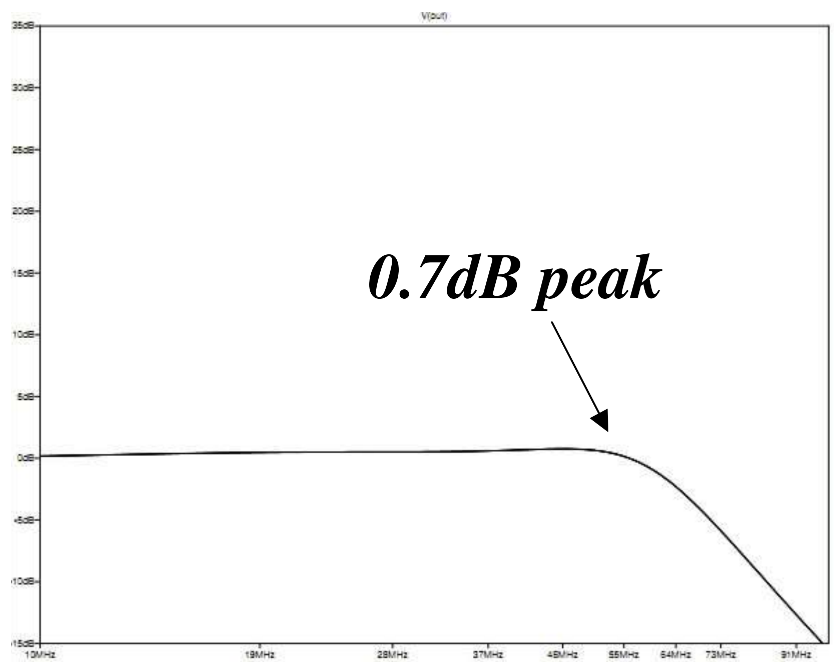
- Familiar circuit has surprising, terrifying but *understandable* behavior:





Cascaded followers: The fix

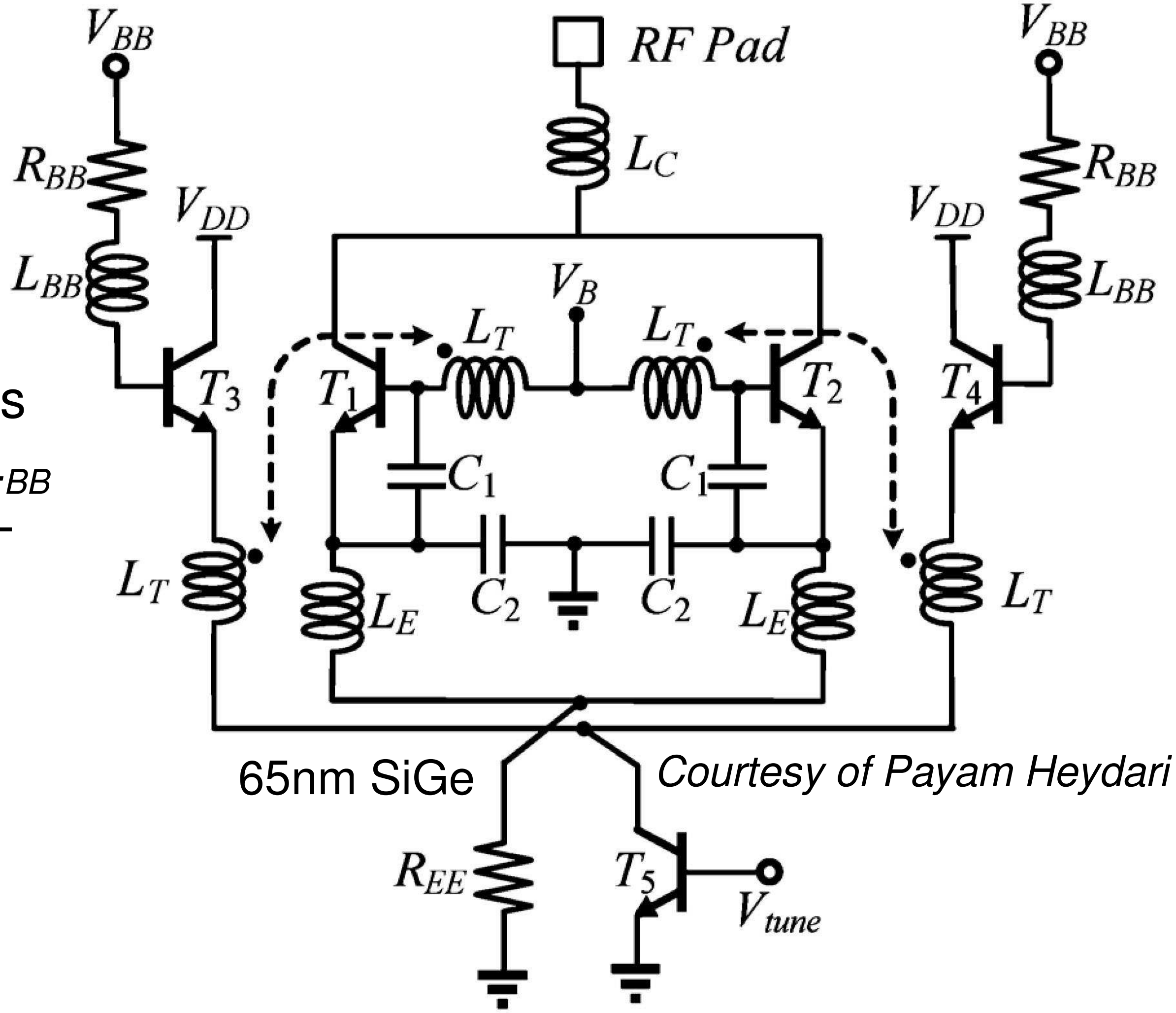
- Problem is negative *R*, so add *R4* (220Ω) to cancel it:



- Shunt R4 with a capacitor to reduce BW loss.

Exploiting SIBS: 200GHz push-push VCO

R_{BB} becomes inductance, L_{BB} becomes Q-enhancing negative R.

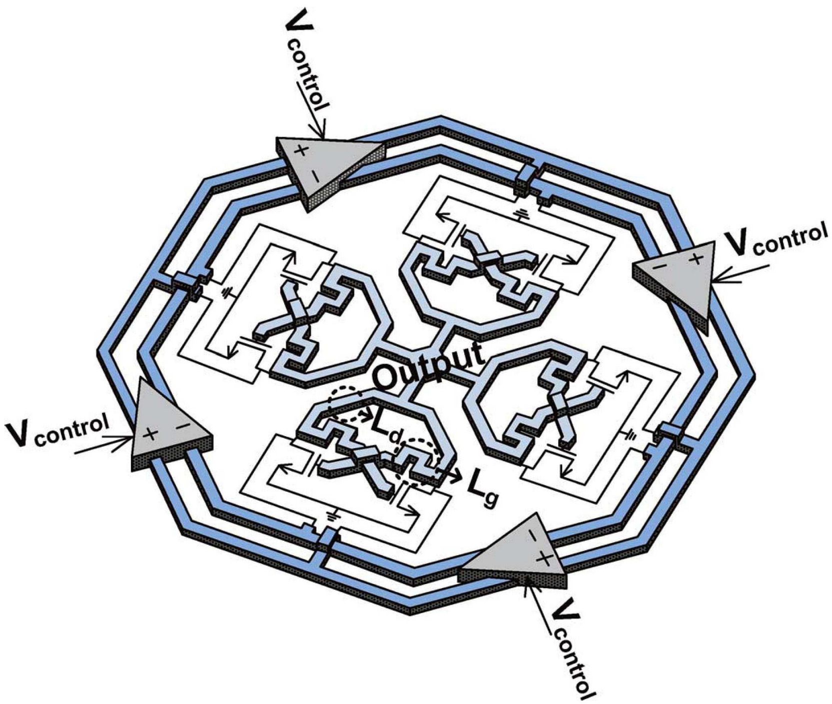


65nm SiGe

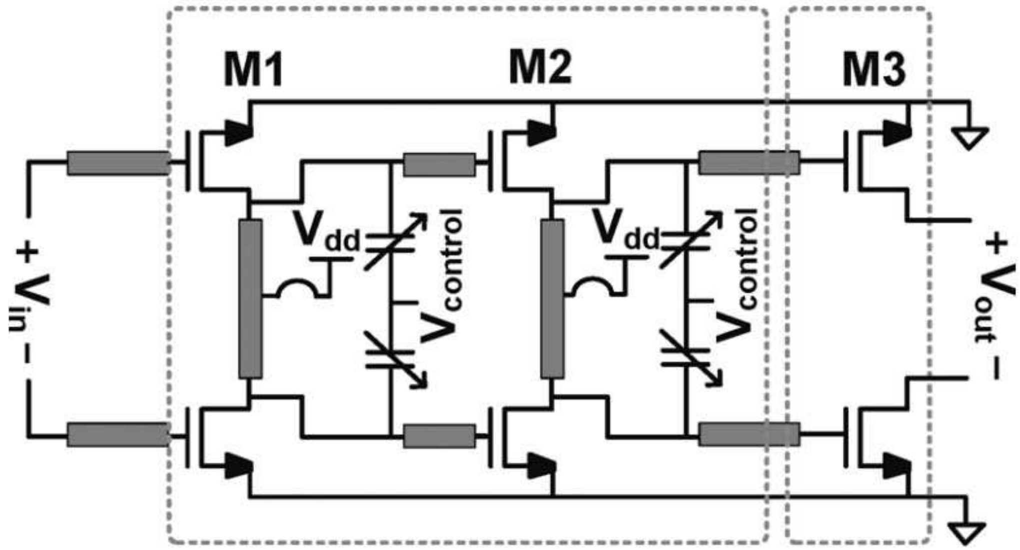
Courtesy of Payam Heydari

Extreme Circuit Design:
Oscillator Tricks

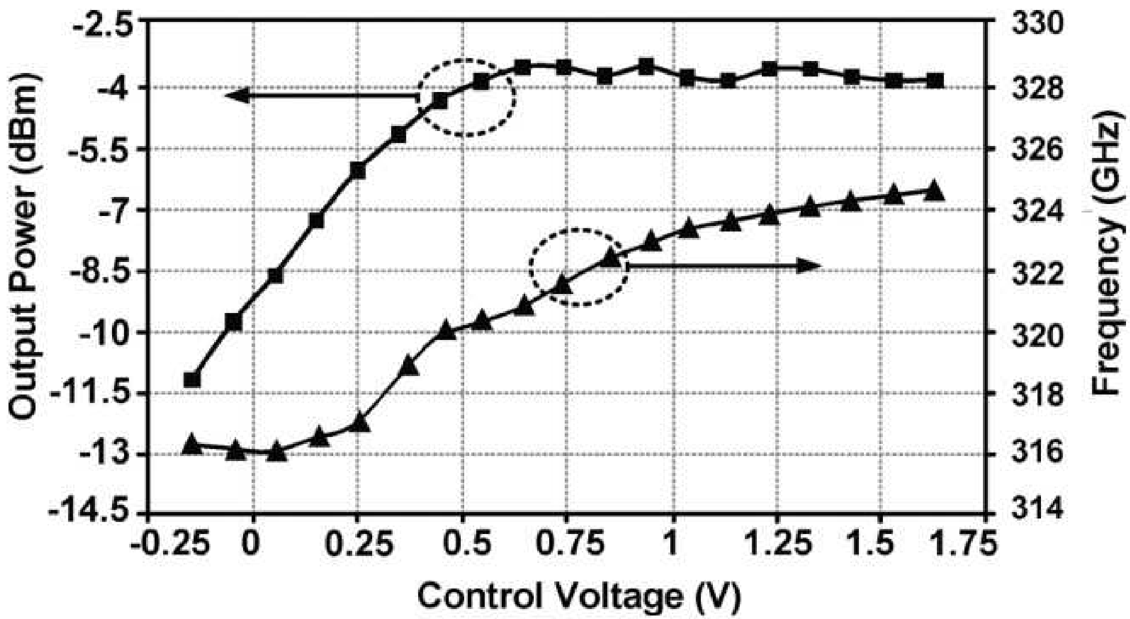
280GHz quad-push oscillator (65nm CMOS)

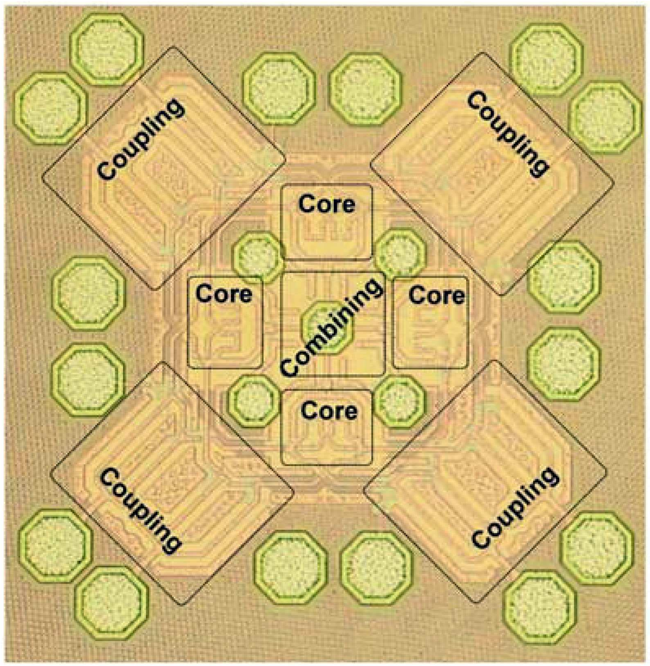


Courtesy of Ehsan Afshari



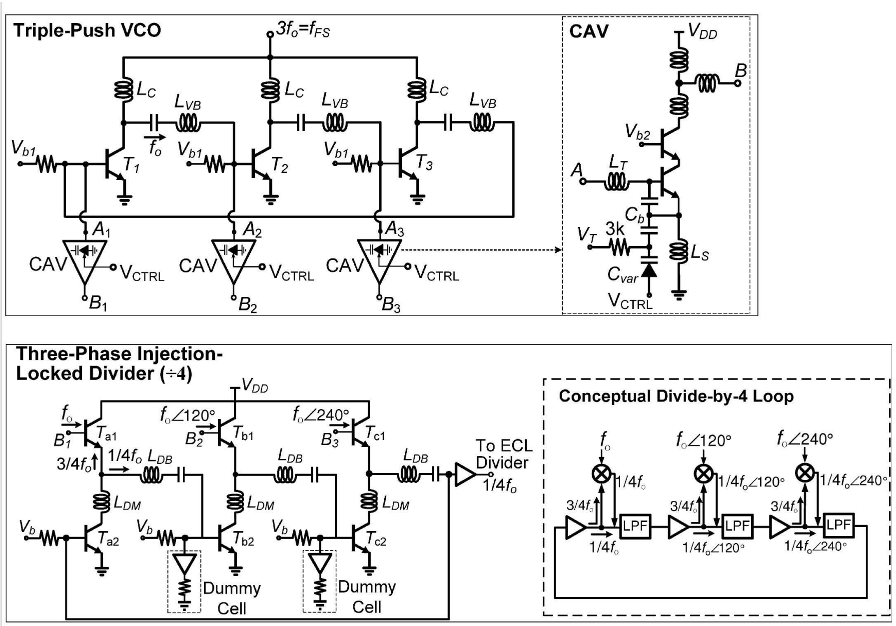
Measured performance of 280GHz oscillator



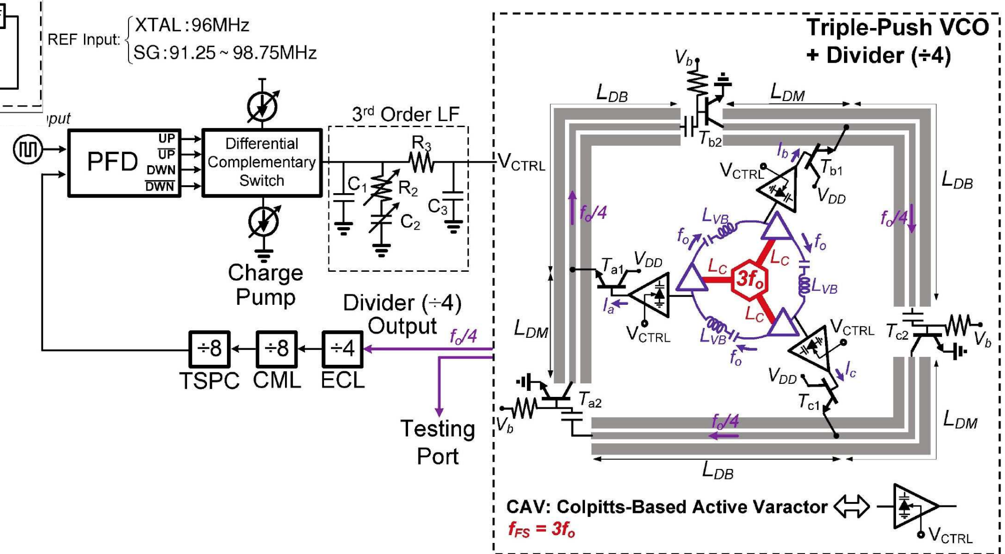


Courtesy of Ehsan Afshari

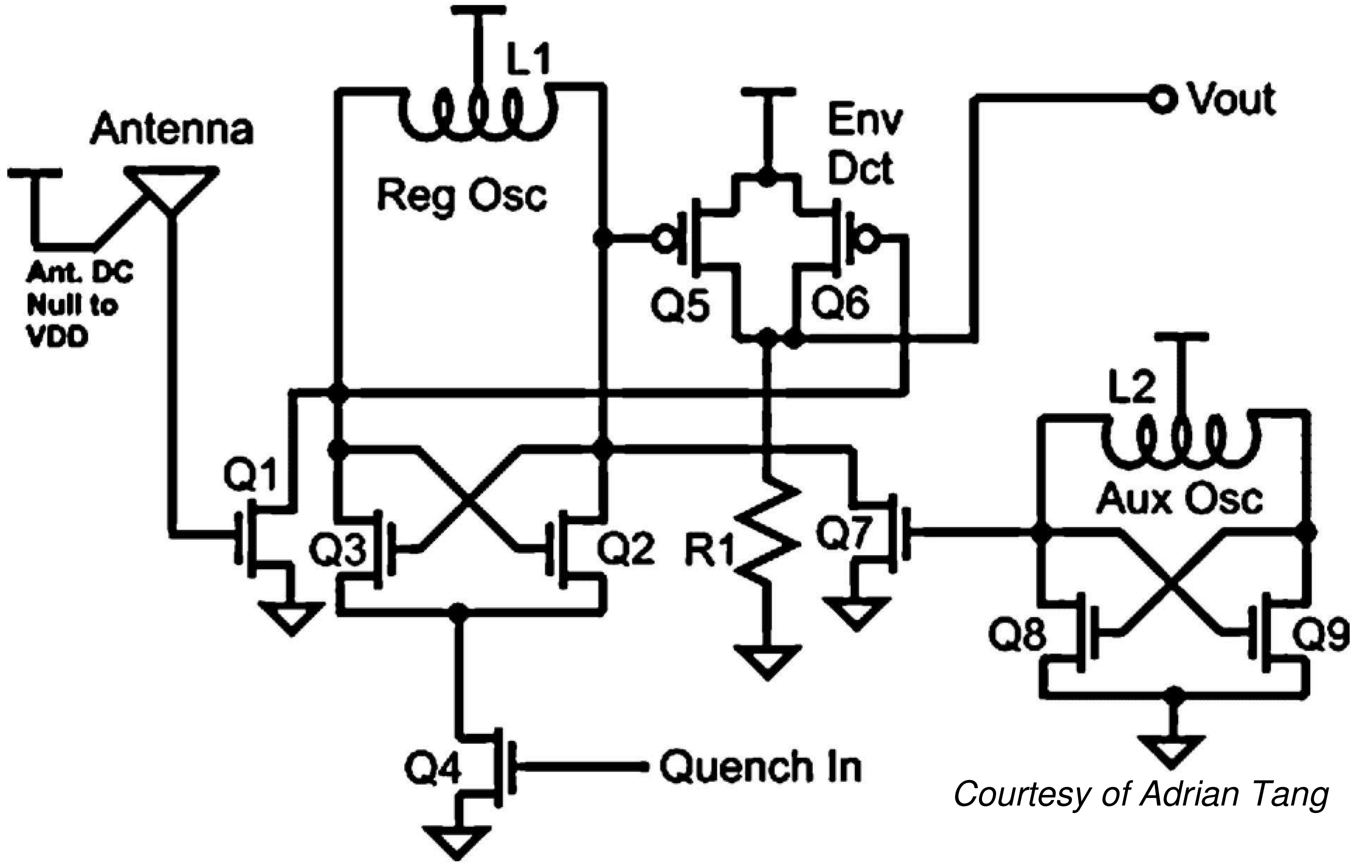
300GHz triple-push SiGe synthesizer



Courtesy of Payam Heydari



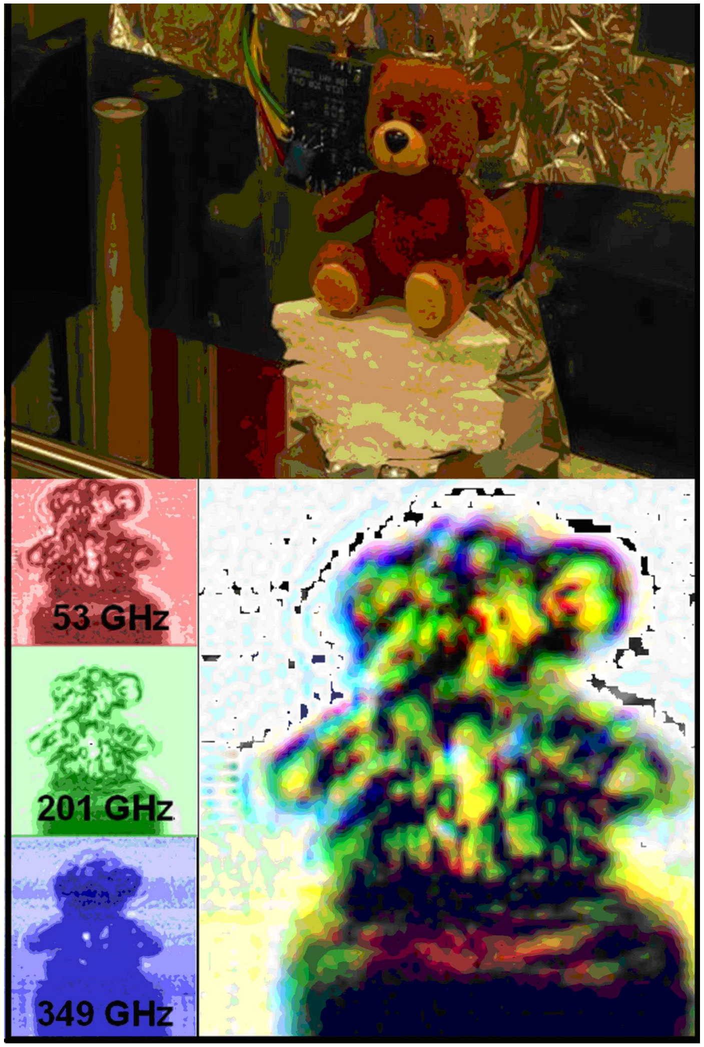
Return of superregeneration



Courtesy of Adrian Tang

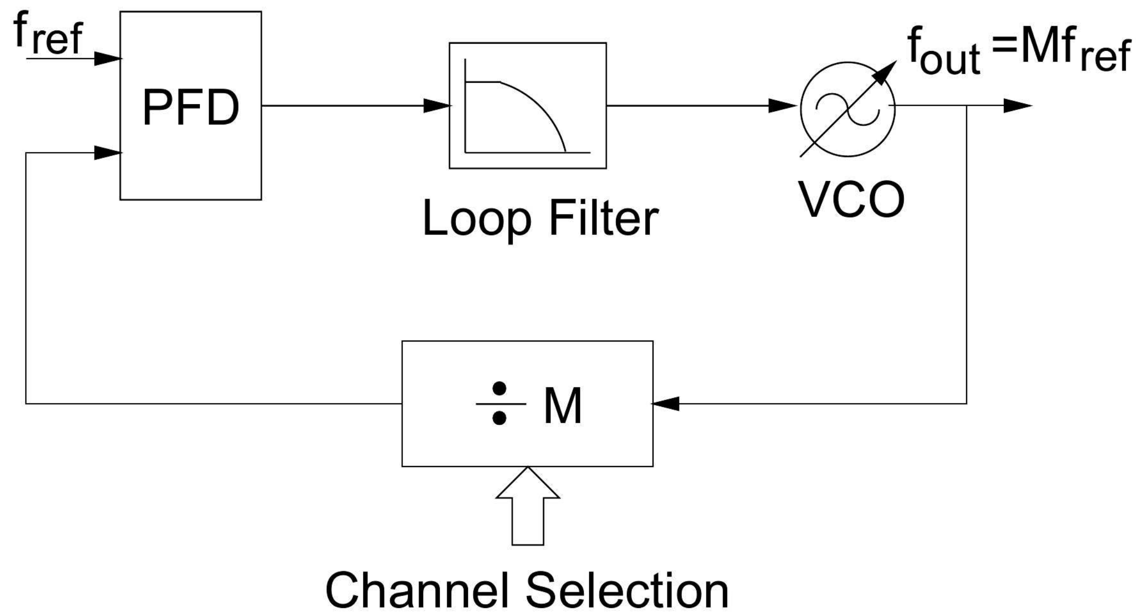
349GHz receiver in 65nm CMOS;
495GHz in 40nm

Tri-band imaging result



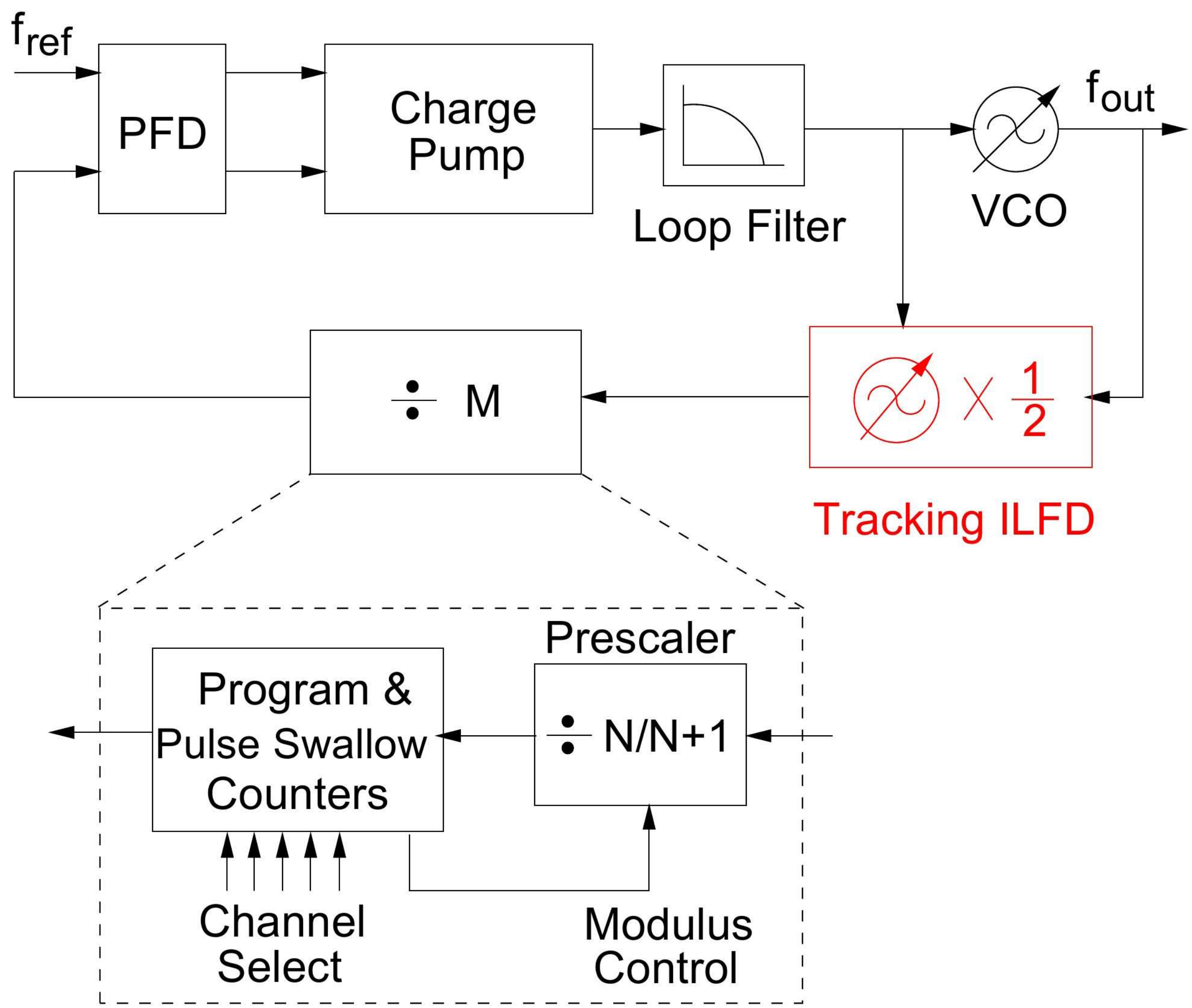
Courtesy of Adrian Tang

PLL frequency synthesis challenges

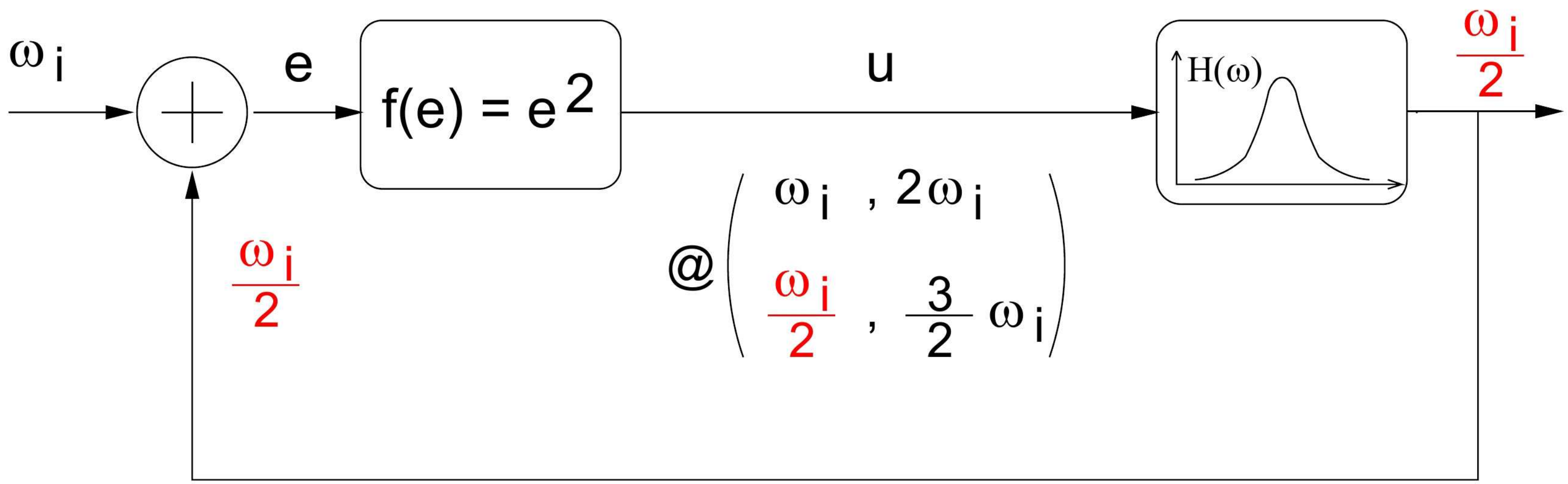


- Reference is a crystal oscillator.
- $f_{out} = \mathrm{M} \times f_{ref}$

Injection-locking eases prescaler design



Injection-locking uses embedded mixing

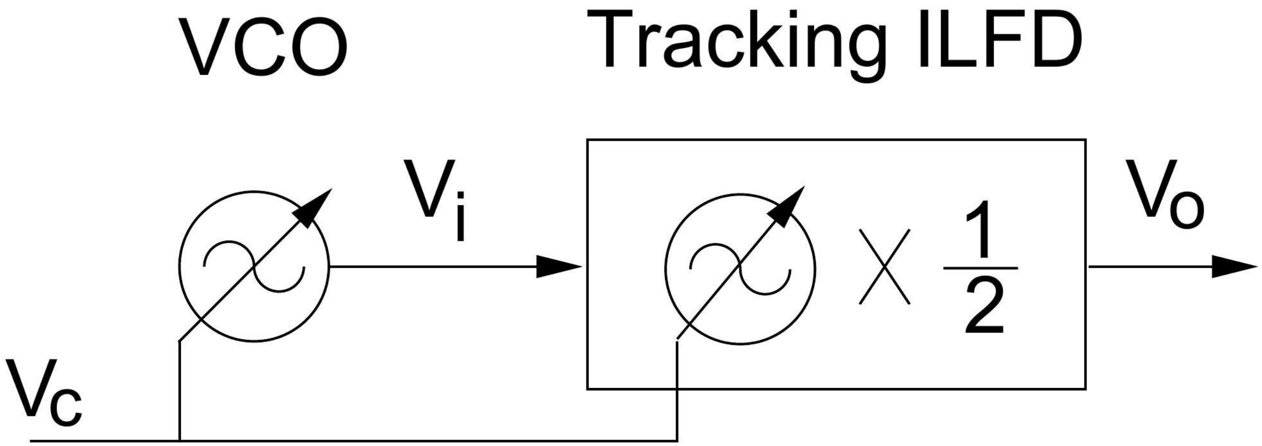


- Example above is a divide-by-2 prescaler.
  - Second-order nonlinearity creates mixing products.
  - Filter preserves only difference component ($\omega_i/2$).

Injection-locked dividers at freq. extremes

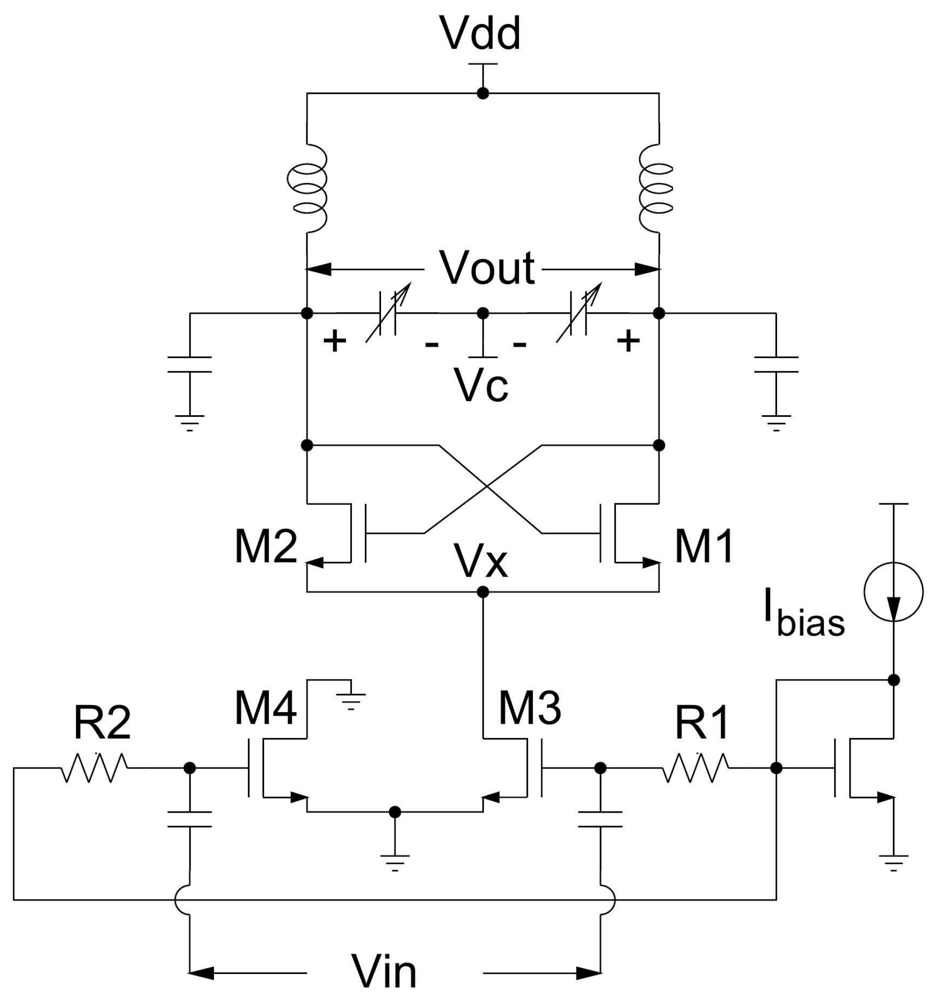
- Maximum input frequency of ILFD can (greatly) exceed device f_T.
  - Key elements are nonlinearity and filtering, neither of which is bounded directly by f_T.
  - Basic requirement is to support oscillation at desired *output* frequency, which is often a tiny fraction of the input frequency.
- Power consumption is also typically much lower than for conventional FF-based dividers.

Injection-locking for frequency synthesis



- Extended locking range provided by tuning ILFD filter to track VCO.

Representative CMOS ILFD

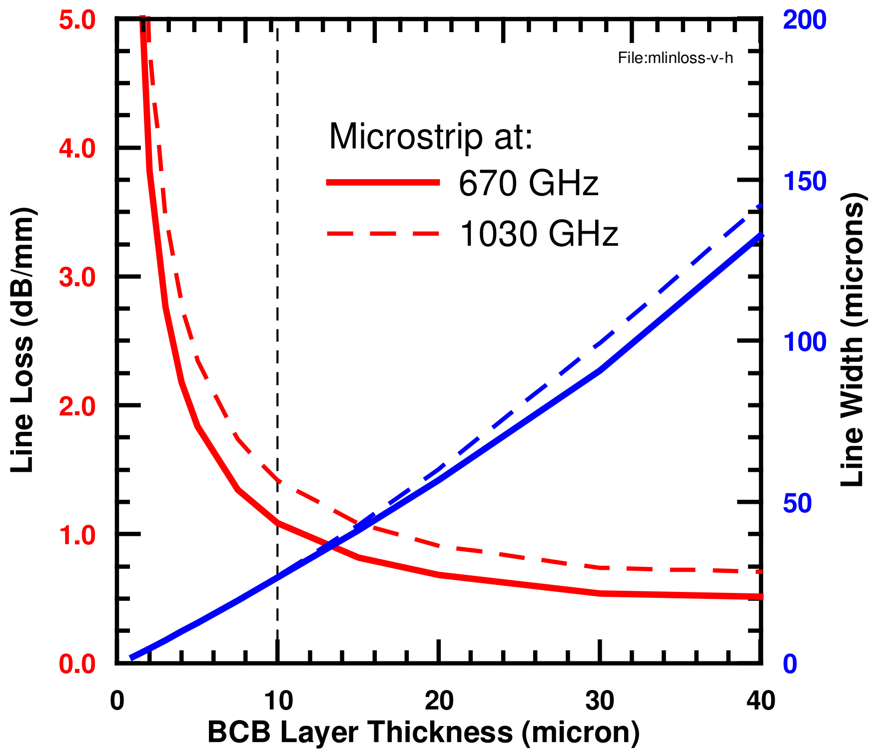


Need interconnect, too

- Mode-free microstrip at 670 GHz and higher requires thin, fragile substrates.
- Primary conductor losses increase as the substrate is thinned according to

$$\alpha_{skin} \propto \varepsilon_r^{1/2} T^{-1}$$

- A low-ε_r substrate such as BCB (ε_r=2.5) reduces attenuation compared to a Si substrate with ε_r of ~11–13.
- 10µm BCB thickness represents good compromise between interconnect density and line attenuation.
    - Loss of the order of 1.5dB/mm at 1THz.



Fixturing interconnect matters, too

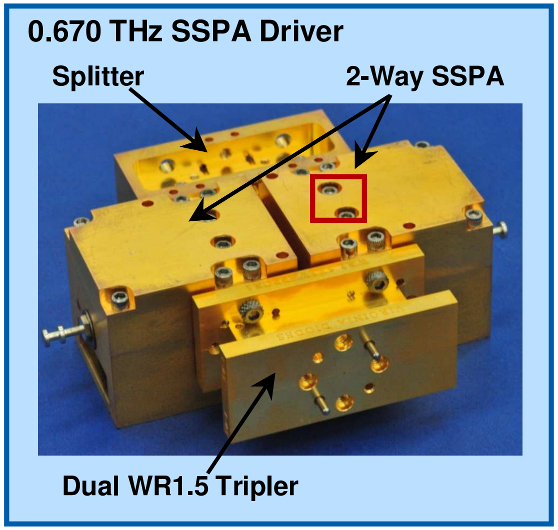
- HE11-mode THz waveguide now available.

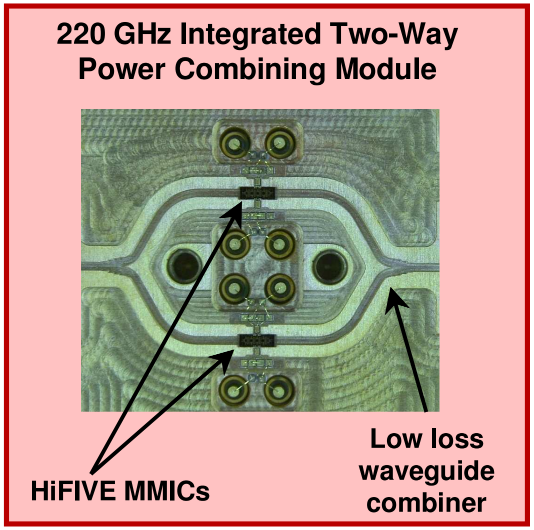
- Attenuation below 3dB/m.

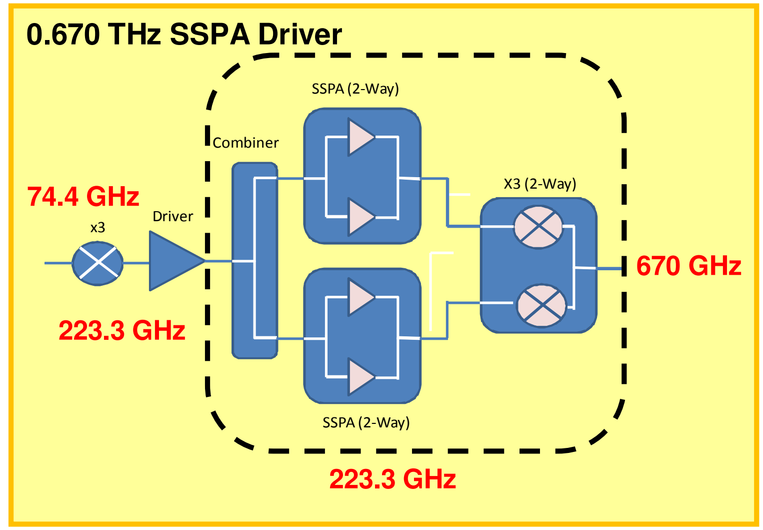
Source: Swissto12

Extreme Circuit Design:
Frequency multiplication

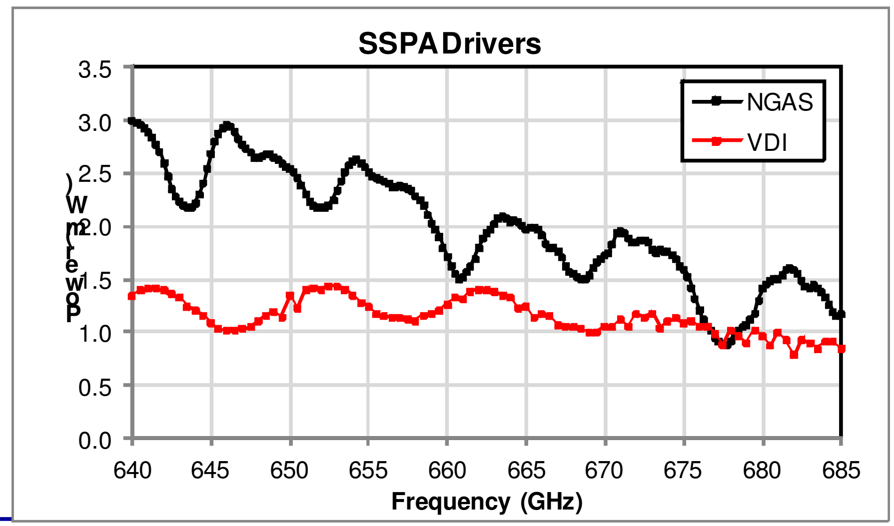
NGC 670 GHz SSPA



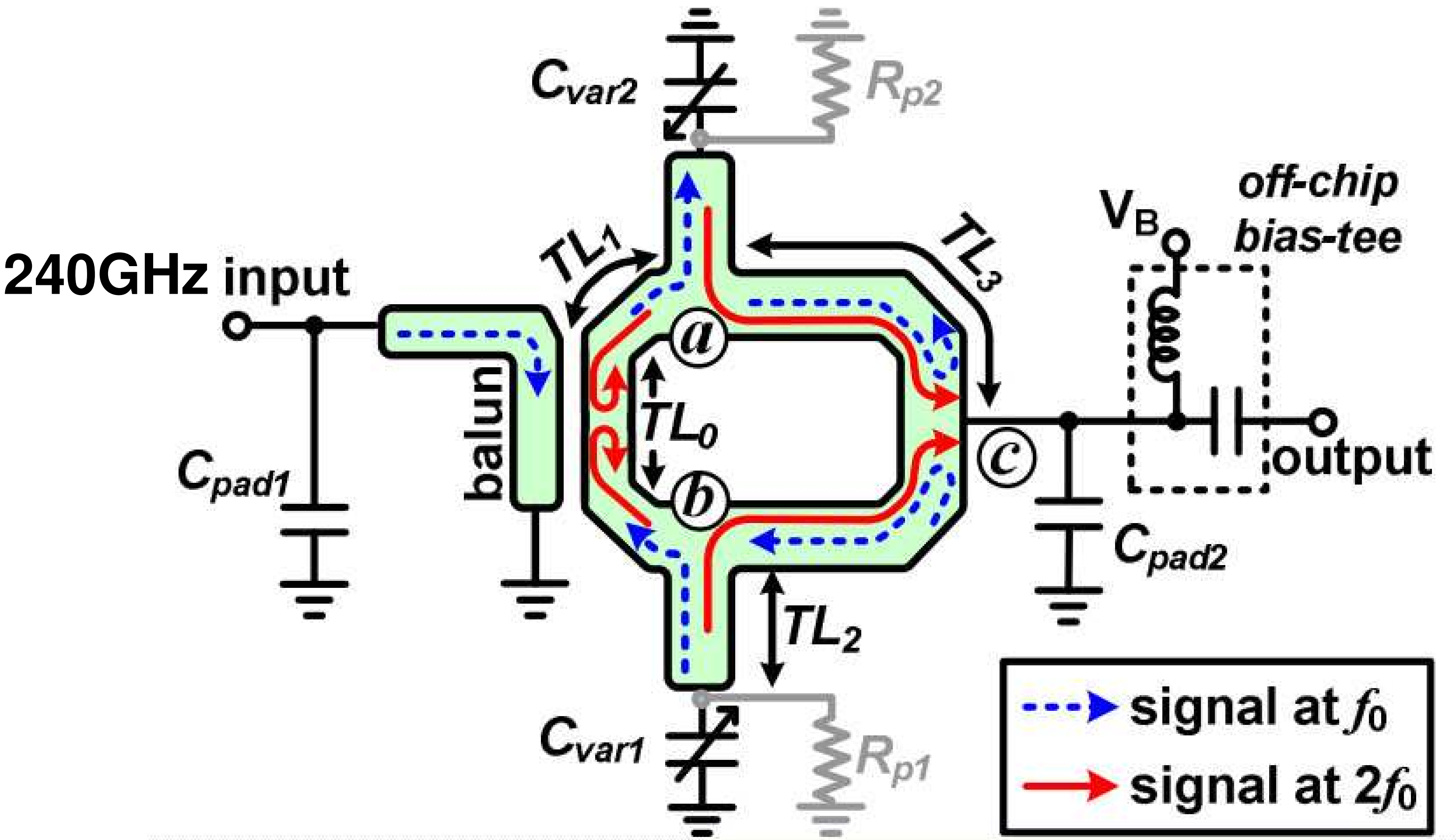




- Power combining of 2 two-way SSPA modules (4-way combining of MMICs in total)
- MMIC housing and waveguide combiner integrated in single module to reduce interconnect losses
- Each SSPA module produces 110 mW @ 220 GHz with 9-13 dB gain
- Tripled, power-combined output is 1-3 mW from 0.640-0.685 THz

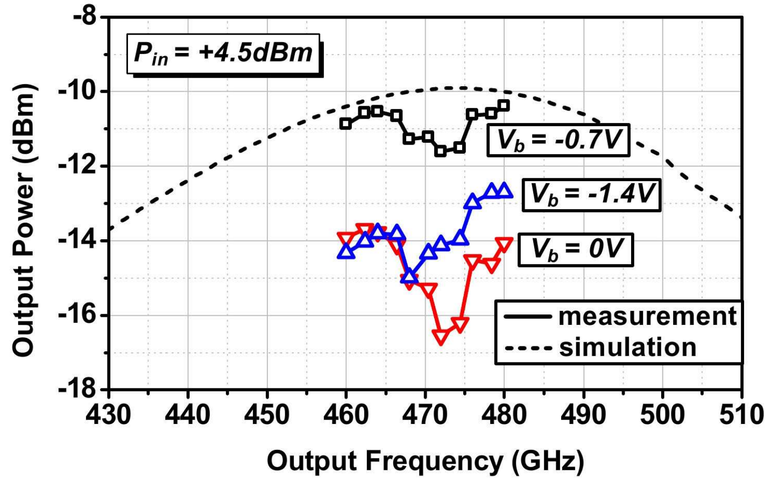


Passive doubler for 480GHz output

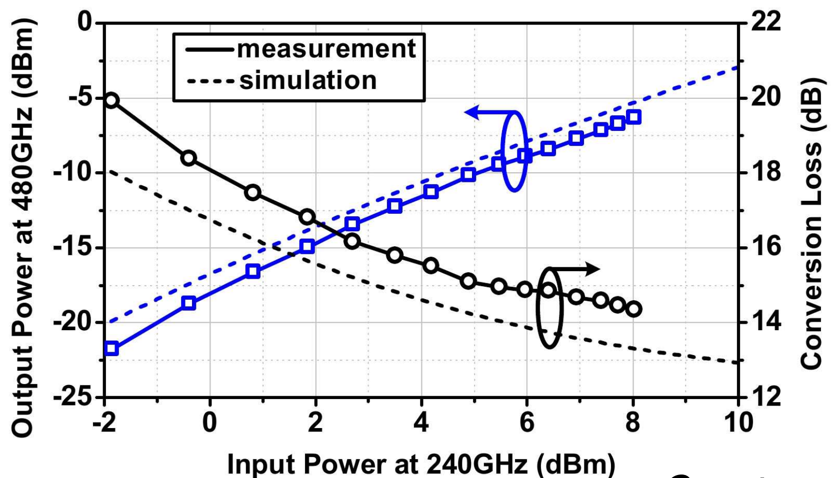


Courtesy of Ehsan Afshari

Measured performance of passive doubler



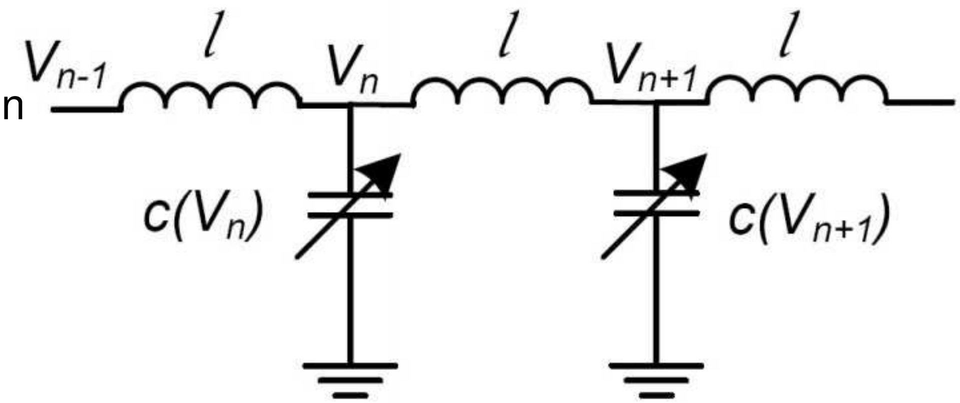
(a)

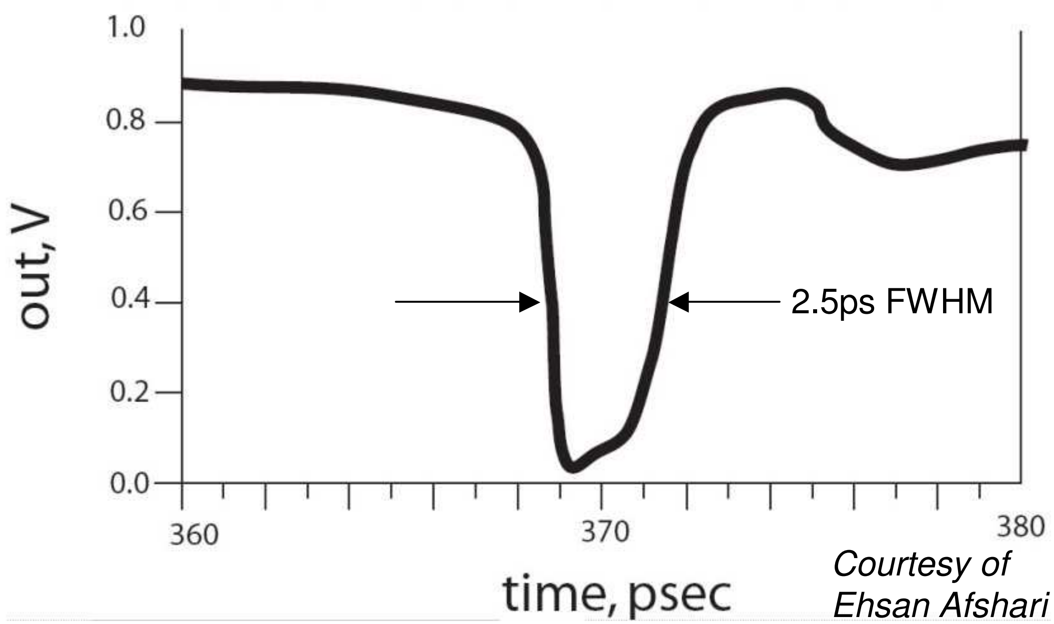


(b)

Courtesy of Ehsan Afshari

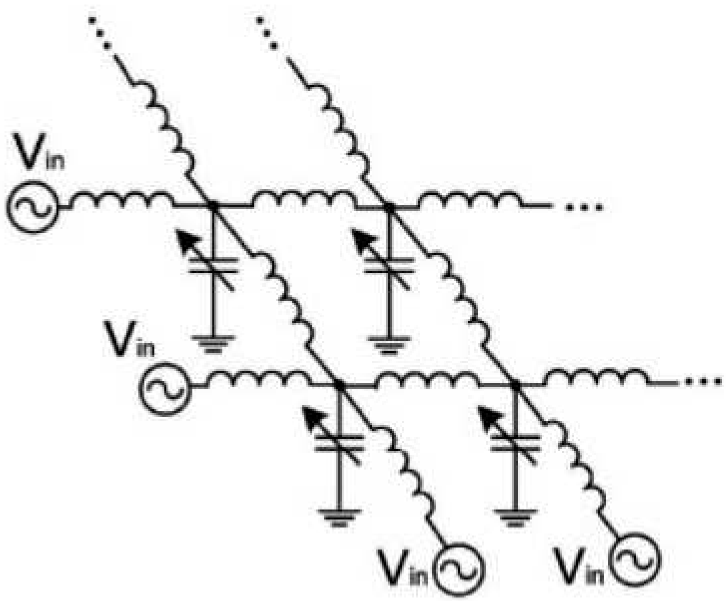
2D shockwave line for fast pulses

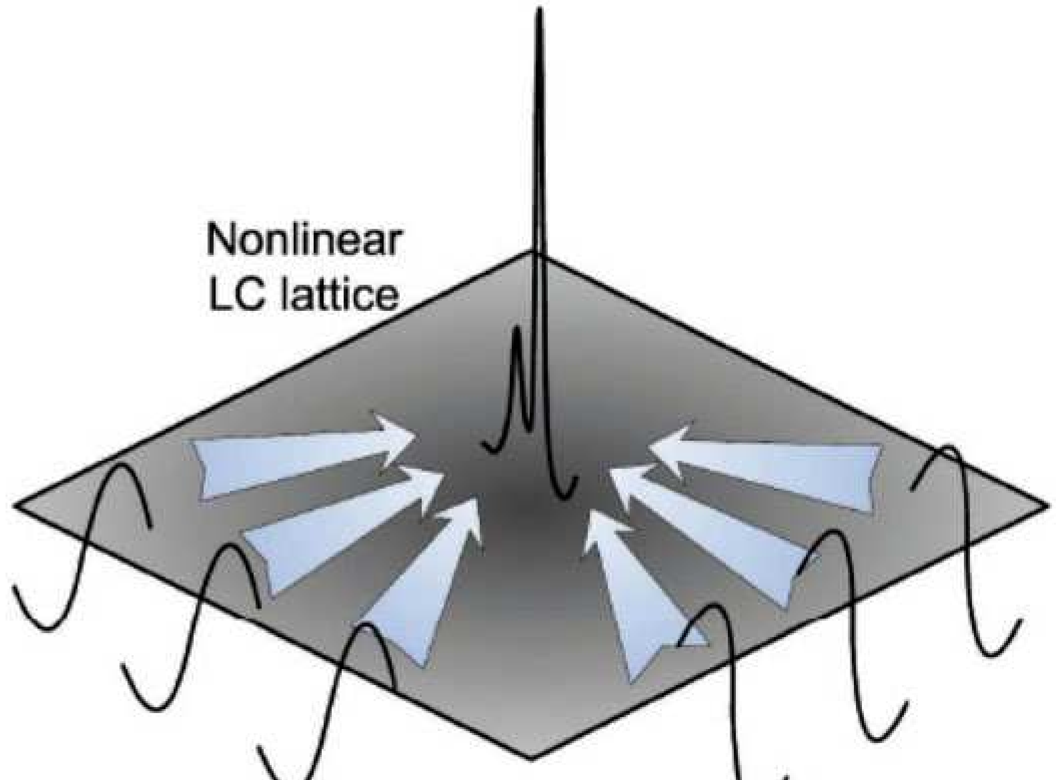




Courtesy of Ehsan Afshari

2D shockwave fabric for higher power





Courtesy of Ehsan Afshari

Summary

- Moore's law continues to give us faster devices.
- Appropriate circuit techniques can allow operation at frequencies near device limits.
- Fuller exploitation of SIBs, nonlinearities and parametric (LTV) phenomena can push frequencies of operation *beyond* f_T.

References

[Goo] J.S. Goo, *High Frequency Noise in CMOS Low-Noise Amplifiers*, Doctoral Dissertation, Stanford University, August 2001.

[Jones] H. E. Jones, US Pat. #3,241,078, "Dual Output Synchronous Detector Utilizing Transistorized Differential Amplifiers," issued March 1966.

[Lee] *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd edition, Cambridge U. Press, 2004.

[Shaeffer] D. Shaeffer and T. Lee, "A 1.5-V, 1.5-GHz CMOS Low *Noise* Amplifier," *IEEE J. Solid-State Circuits*, v.32, pp. 745-758, 1997.

[Terrovitis] M. T. Terrovitis and R. G. Meyer, "Noise in Current-Commutating CMOS Mixers," *IEEE Journal of Solid-State Circuits,* vol. 34, No. 6, June 1999.